The Effect of Area Level Deprivation on Obesity in New Zealand: Analysis of The New Zealand Health Surveys

---

A thesis submitted in partial fulfilment of the requirements
for the Degree of Master of Health Sciences
in the University of Canterbury

by Andrew Adiguna Halim (ID: 62376311)
University of Canterbury

July 30, 2017

Contents

List of Tables

List of Figures

Listings

Acknowledgements

I wish to express sincere gratitude to Dr. Arindam Basu and Prof. Ray Kirk for their assistance in the preparation of this manuscript. I enjoyed every bit of my research and writing process because of their kind support and guidance. Not only that, they have also taught me how to critically formulate my research arguments and what it means to do research. Though the process has not been easy, this unforgettable journey allows me to wholeheartedly appreciate both the challenges and the joy of doing research even more.

I am also grateful to Statistics New Zealand for their assistance in providing me with the New Zealand Health Survey data. They also had assisted me many times with the questions I had regarding the data.

I would also like to express special thanks to Pat Coope whose assistance in statistics has taught me to better understand the importance of statistical tests and its implication in interpreting my research findings. I also want to thank Daniel Gerhard and Elena Moltchanova for setting up statistics workshops and introducing me to Bayesian inference. The knowledge and experience I gained helped me see the strange and peculiar things around the world with better clarity and comprehension.

Special thanks go to Robyn Johnston, Chris Chen, Lim Su Hui, Wade Stent, Kate Reid, Gopal Panta, Uzma Irfan, Llyween Cooper, Helen Mataiti, Khoa Vo, Arun Blasi, Babajide Onademuren, Mark Owusu, Vikram Lakshminarayanan, Ashleigh Sandy Gozali, and Anne Whitfield. I am really thankful for them, listening to their stories and experiences never fail to cheer me up. Their practical advice on daily life also had helped me on multiple occasions when I had no clue what I should do. Though it may seem trivial, the moral support, the jokes, and the laugh are the things that keep me going and make me smile.

Finally, I must express my very profound gratitude to my parents, my brother, my best friends in Indonesia for providing me with unfailing support throughout my years of study. This accomplishment would not have been possible without them. Thank you.

Abstract

Background: In New Zealand 31% of adults and one in nine children aged 2–14 years were obese in 2014/15. According to the 2014/15 New Zealand health survey, people who lived in the most deprived areas were almost two times more likely to be obese than those who lived in the least deprived areas.

Objectives: To assess the association between the area level deprivation and overweight/obesity prevalence while controlling for demographic variables and health-related behaviours.

Method: The analyses were based on the health surveys conducted from 2002/03 to 2014/15 in New Zealand. This study assessed area level deprivation using the New Zealand Deprivation Index. Proportional odds logistic regression with sampling weights was used to answer the objective.

Results: Adults who lived in deprivation quintile five were more likely (OR 1.46, 95% CI 1.33–1.60) to be in the higher Body Mass Index (BMI) tertiles compared with those who lived in quintile one, and the association was stronger in children (OR 1.76, 95% CI 1.50–2.06). Pacific (OR 8.41, 95% CI 7.22–9.79) and Māori adults (OR 3.01, 95% CI 2.75–3.29) had a higher odds to be in the higher BMI tertiles compared with Europeans.

Conclusion: Based on the representative nationwide dataset, I found that higher deprivation quintiles were significantly associated with a higher likelihood of being in the higher BMI categories after controlling for demographic information and health behaviours. Public health interventions targeting obesity should take into account the deprivation level and ethnic groups' composition of the communities.

Glossary

agnostic A position/attitude without prior beliefs about the subject. 9

AHEAD Action for Health in Diabetes. A multi-centre, randomised controlled trial, designed to determine whether intentional weight loss reduces cardiovascular morbidity and mortality in overweight individuals with type 2 diabetes. 7

anastomotic rupture A break in the surgically co-joined structure of the body. In this research context, it relates to the intestines or stomach. 8

apoptosis A regulated cell suicide process, usually triggered when the cells are no longer needed or posing a threat to the organism. 9

AUDIT Alcohol Use Disorders Identification Test (AUDIT) is a 10-item screening tool developed by the WHO to assess alcohol consumption, drinking behaviors, and alcohol-related problems. A score of 8 or more is considered to indicate hazardous or harmful alcohol use. 22, 28

bariatric surgery Surgical procedures on obese people to reduce their weight. Examples of such procedures are: reducing the size of the stomach with gastric band, removing a portion of the stomach, re-routing the small intestines to a small stomach pouch, etc. 2, 8

BMI Body mass index (BMI) is a person's weight in kilograms divided by the square of height in metre. vii, xi, 1, 2, 4–6, 8, 9, 12, 13, 15, 16, 21, 23, 26–29, 34, 35, 39–45, 47

BMI-H BMI-Heritability is the heritability coefficient of BMI. A proportion of the variation in BMI that can be explained by genetics. 8

Bogalusa Heart Study The Bogalusa Heart Study is a biracial (65% white and 35% black) community-based, long-term investigation of the early natural history of CVD beginning in childhood. 5

CDC Centers for Disease Control and Prevention. 5

CNS Central nervous system. A part of the nervous system consisting of the brain and spinal cord. 2, 9

comorbidities The presence of one or more medical conditions associated with the primary disorder. In this research context, any medical conditions associated with obesity such as type II diabetes melitus, hypertension, dyslipidemia, etc.. 1

CVD Cardiovascular disease. xi, xii, 5

DALY Disability-Adjusted Life Years (DALY) is a measure of overall disease burden, expressed as the number of years lost due to disability or early death. 4

doubly-labelled water A method that measures metabolic rate using an uncommon isotope of water (deuterium and oxygen-18) excreted from the body. 13

FHR The Framingham Heart Study (FHS) is a large cohort study in the town of Framingham, Massachusetts. It has led to the identification of major CVD risk factors and many others. 4

gastrointestinal Relating to the stomach and intestines. 4, 8

GDP Gross Domestic Product (GDP) is a monetary measure of the market value of all final goods and services produced in a given period (usually yearly) in a country. Nominal GDP estimates are commonly used to determine the economic performance of a country and to make international comparisons. 8

GNI Gross national income (GNI) is the total domestic and foreign output claimed by residents of a country. It consists of gross domestic product plus factor incomes earned by foreign residents, minus income earned in the domestic economy by non-residents. 11

GWAS Genome-Wide Association Study (GWAS) is an examination of many common genetic variants in different individuals to see if any variant is associated with a trait. 2, 9

HDL High-density lipoprotein (HDL) is one of the five major groups of lipoproteion. Increasing concentrations of HDL particles are strongly associated with decreasing accumulation of atherosclerosis within the walls of arteries. 5, 7

hemorrhage An escape of blood from a ruptured blood vessel. 8

HHID Household ID. It is a household identifier for the respondents of the NZHS. 24

HR Hazard ratio (HR) is the probability of an event (death or disease) occurring in the study group compared to the probability in the comparator group at any time point. 4, 5

IOTF International Obesity Task Force (IOTF) is an organisation designed to combat obesity around the world. Later known as World Obesity Federation. xiii, 1, 5, 21, 23

LDL Low-density lipoprotein (LDL) is one of the five major groups of lipoprotein and a risk marker for cardiovascular disease. Higher value of LDL translates to higher risk of cardiovascular disease. 5

NHANES The National Health and Nutrition Examination Survey (NHANES) is a survey that is designed to assess the health and nutritional status of adults and children in the United States. 7

nicotinic cholinergic receptors They are receptors that respond to the neurotransmitter acetylcholine. 16

NZHS New Zealand Health Survey. 1, 3, 15, 19, 20, 23–26, 45, 46

obstructive sleep apnea A medical condition where the breathing stop and start repeatedly during sleep caused by an obstruction in the airway. 8

OECD The Organisation for Economic Co-operation and Development (OECD) is an organisation that aims to promote policies that will improve the economic and social well-being of people around the world. 1, 40, 41

price elasticity A measure of the effect of a price change or a change in the quantity supplied on the demand for a product or service. 12

SES Socio-Economic Status. 3

SNP A Single-nucleotide polymorphism (SNP) is a variation in a single nucleotide that occurs at a specific position in the genome. For example, a SNP may replace the nucleotide cytosine (C) with the nucleotide thymine (T) in a certain location of DNA. 2, 9

SSB Sugar-sweetened beverage. 13

triglyceride It is an ester derived from glycerol and three fatty acids. A high concentration of triglyceride is associated with an increased risk of heart diseases, diabetes, and stroke. 5, 7

WC Waist Circumference. 1, 5

WHO World Health Organization. 5, 22, 23

WHR Waist/hip ratio. 1, 5

World Obesity Federation It is an organisation which has a mission is to lead and drive global efforts to reduce, prevent and treat obesity. Formerly known as IOTF. xii, 5, 21

1 Background

Obesity is a chronic condition associated with an increased risk of hypertension, dyslipidaemia, type II diabetes, stroke, congestive heart failure, coronary artery disease, increasing health costs, and many other comorbidities [Guh et al., 2009, Jensen et al., 2013, Barness et al., 2007]. Those chronic complications require resources to treat and in turn has led to the growing healthcare expenditure in the population. An individual with a body mass index (BMI) between 30 and 35 is estimated to require 19% more health expenses compared with those with normal BMI, and as the BMI breaks through 35 they would need 51% more [Buchmueller and Johar, 2015]. The rate of obesity around the world has doubled since 1980; with 39% and 13% of adults aged 18 and over, were overweight and obese respectively in 2014 [World Health Organization, 2015]. Among The Organisation for Economic Co-operation and Development (OECD) countries, New Zealand was ranked third in its adult obesity prevalence in 2017 [OECD, 2017]. According to the New Zealand Health Survey (NZHS) in 2014/15, 31% of adults and one in nine children aged 2–14 years were obese [Ministry of Health, 2015a]; the rates were higher among Māori, Pacific, and those with the highest level of socio-economic deprivation [Ministry of Health, 2015a, Swinburn et al., 2014]. From New Zealand's economic perspective, the estimated direct cost of treating obesity is NZ$460 million in 2004, while the indirect cost is NZ$370 million in the same year [Ministry of Health, 2009]. The actual direct cost calculated in 2006 was NZ$624 million or 4.4% of New Zealand′s total health care expenditure, in addition to NZ$225 million due to lost productivity [Swinburn et al., 2014].

The criteria for overweight in adults is defined by a BMI between 25 and 30, and obesity is defined by a BMI of 30 or over [Centers for Disease Control and Prevention, 2016]. Whereas the criteria for defining overweight and obesity for those aged between two and below 18 years are based on the International Obesity Task Force (IOTF) BMI-for-age chart. It is defined in such a way so that it will merge with the adult's BMI criteria once they reach 18 years of age [Cole and Lobstein, 2012]. Other adiposity measurements such as waist circumference (WC) and waist/hip ratio (WHR) were thought to have superiority in predicting health risks associated with obesity. Nevertheless, a meta-analysis comparing BMI, WC, and WHR in predicting the incidence of diabetes found no significant difference [Vazquez et al., 2007]. For one standard deviation increase in BMI, WC, and WHR; the age-adjusted odds ratio for diabetes in men (women) were 1.52 (1.59), 1.54 (1.70), and 1.53 (1.50). For hypertension, they were 1.68 (1.55), 1.66 (1.51), and 1.45 (1.28) [Huxley et al., 2010]. Considering the high correlations across adiposity measurements and none showing superior discriminatory capability, it is clinically irrelevant to choose one measurement over the other [Vazquez et al., 2007, Huxley et al., 2010].

Individual management options for treating obese patients range from lifestyle interventions to invasive surgical procedures. Lifestyle interventions aim to improve physical activity levels and dietary habits through various training regiments and diet restrictions [Jensen et al., 2013]. However, such interventions only produce slight weight loss [Shaw et al., 2006], have a poor long-term adherence rate [Sheperd, 2005, Kovacs et al., 2014] and a high rate of regaining weight loss after stopping the intervention [de Vos et al., 2016]. On the other hand, bariatric surgery was able to achieve a long lasting weight loss and a significant reduction in cardiovascular events compared with those receiving non-invasive treatments [Sjöström et al., 2012]. Furthermore, there is a high rate of resolution of chronic conditions such as diabetes, hypertension and obstructive sleep apnea following bariatric surgery. Nevertheless, bariatric surgery is not only expensive but also has an operative mortality rate of 0.1–1.1% [Buchwald et al., 2004]. In addition, only those who are morbidly obese are eligible to undergo the surgery [Neff et al., 2013, Sheperd, 2005, Jull et al., 2011]. With the current available individual interventions, it does not appear as though these will be sufficient to curb the obesity epidemic at a population level.

Since the conception and increasing use of Genome-Wide Association Studies (GWAS), more than 20 loci have been found to have an association with obesity. Several of these genes are expressed in the Central Nervous System (CNS) which suggests involvement in energy homeostasis through feeding regulation [Herrera and Lindgren, 2010]. Nevertheless, two commonly identified variants of Single-Nucleotide Polymorphisms (SNPs) that have been linked to obesity, in an intron of FTO gene (situated in chromosome 16) and within a presumed regularly site 188kb downstream of the MC4R gene (situated in chromosome 18), only account for less than 2% variance in adult BMI. GWAS on obesity genetics may be able to shed more insights on the phatophysiology of obesity. Unfortunately, it will not be feasible anytime soon to find the genetic solution to tackle the current obesity epidemic at a population level [Bogardus, 2009].

Urbanisation has changed how people live in relation to their social as well as physical environment [Malik et al., 2013b]. The compact urban environment draws food outlets that provide cheap and unhealthy meals closer to homes [Kearney, 2010]. Advancement in technology has also increased the demand for mechanised works and activities, which in turn decrease the physical activity levels of a population. On top of that, people spend more of their leisure time watching television or browsing the Internet [Malik et al., 2013b]. Changes in agricultural practice in the past 50 years have also made our food supply bountiful and cheap [Kearney, 2010, Johnson and Wardle, 2011]. The world has increased its food consumption by 400 kcal per person per day between 1969/1971 and 2005/2007. Types of diets also shifted from staples (e.g., roots and tubers) into livestock products and vegetable oils, which are denser in energy [Alexandratos and Bruinsma, 2012, Kearney, 2010]. Food scarcity is no longer an issue in high-income countries with a strong welfare system [Dinsa et al., 2012]. Instead, the problem has shifted to an unhealthy living environment that promotes positive energy balance. Positive energy balance, either surplus of energy intake (e.g. energy-dense diet) or low energy output (e.g. sedentary lifestyle), can contribute to obesity [Giskes et al., 2011, Malik et al., 2013b, Kelly and Swinburn, 2015].

Disparities in the incidence of obesity exist across people with different socio-economic statuses (SES). Dinsa, et al. (2012) conducted a systematic review looking at the association between socio-economic status and obesity in low- and middle-income countries (defined by the World Bank as countries with an income per capita up to US$12,275). Obesity is prevalent among high SES groups in low-income countries. Poor families in low-income countries still experience food scarcity and engage in manual labour; those factors protect them from obesity. Whereas, richer families tend to be overweight/obese because they have access to surplus food [Dinsa et al., 2012] and excess weight can be perceived as a sign of wealth [Wu et al., 2015]. As the income level of a country increases, the relationship between SES and obesity becomes less clear. This may be explained by less apparent food shortages even among poor families in middle-income countries as these countries have more resources to help the poor [Wu et al., 2015, Dinsa et al., 2012]. In contrast, low SES is associated with higher risk of overweight/obesity in high-income countries [Wu et al., 2015, Newton et al., 2017]. Poor families in these countries consume low-cost energy-dense food and have less opportunities to participate in physical activities while their well-off counterparts have more resources to choose healthier diet and lifestyle [Wu et al., 2015]. These tell us that the association between SES and obesity is not straight-forward, and the direction of the association changes depending on the country's income level.

In New Zealand, according to the 2014/15 health survey, people who lived in the most deprived areas were twice as likely to be obese than those who lived in the least deprived areas [Ministry of Health, 2015a]. Other studies in the United States (US) [Eagle et al., 2012], the United Kingdom (UK) [Stafford et al., 2010], and Sweden [Li et al., 2014] also support the finding that people who live in a deprived neighbourhood are more likely to be obese. In addition, the incidence of type II diabetes is higher in more deprived areas [Evans et al., 2000, Maier et al., 2013]. These suggest there may be underlying processes of socio-economic indicators affecting the incidence of obesity and diabetes. To my knowledge, there is no study in New Zealand that assesses the association between socio-economic deprivation and obesity. This study will assess the association between area level deprivation and overweight/obesity prevalence using individual data from the NZHS in 2002/03 to 2014/15. Relevant demographic information and behaviours that may influence obesity prevalence will be controlled.

2 Literature Review

2.1 Health Risks Associated with Obesity

Obesity is a growing problem in New Zealand. Not only it is a major contributor to Disability-Adjusted Life Years (DALY) lost but obese people also are associated with an increased risk of developing chronic conditions [Ministry of Health, 2009]. Chronic conditions such as hypertension and diabetes mellitus type II require long term management and once developed, they are irreversible and require a lot of resources to manage. Additionally, obesity is associated with a higher risk of cardiovascular diseases, pre-eclampsia, respiratory problems, depression, infertility, gastrointestinal diseases, cancers, and a lower life expectancy [Banjare and Bhalerao, 2016]. From a biological perspective, obesity is known to cause insulin resistance, endothelial dysfunction, and promote atherosclerosis formation. All of these lead to the development of chronic non-communicable diseases typically found in obese people [Antonini et al., 2007]. The pathophysiology of obesity-associated chronic conditions is outlined in Figure 2.1.

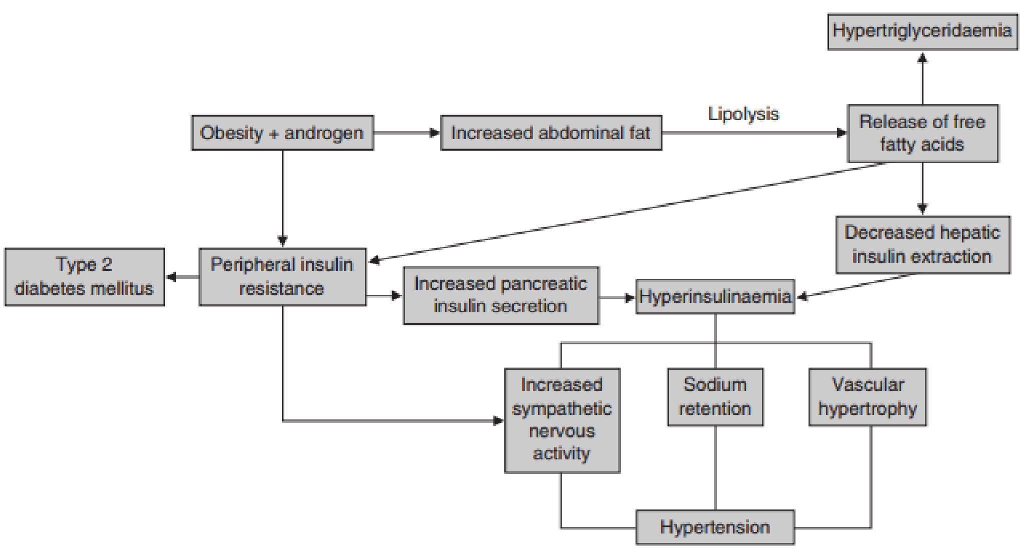

Figure 2.1: Pathophysiological Pathway of Obesity-Associated Chronic Conditions [Antonini et al., 2007]

The risk of obesity-associated chronic diseases is thought to positively associate with the duration of obesity. A study using the data from the Frammingham Heart Study (FHR) have found that obesity duration was associated with the risk of developing type II diabetes. the Hazard Ratio (HR) for type II diabetes mellitus was 1.11 for men and 1.06 for women per additional 2-year increase in the duration of obesity [Abdullah et al., 2011]. However, the association between obesity duration and type II diabetes was heavily attenuated when time-varying BMI and current BMI were included in the analysis. The HR for type II diabetes after the adjustment was 1.02 per additional 2-year increase in the duration of obesity [Hu et al., 2014]. Even though both current BMI and duration of obesity are significant, current BMI appears to exert a stronger influence in predicting the risk of developing type II diabetes.

The health risk associated with obesity in early age is carried to adulthood. Obese children and adolescents have higher fasting glucose, fasting insulin, triglyceride, systolic blood pressure, prevalence of impaired glucose tolerance and lower high-density lipoprotein (HDL) cholesterol compared with their non-obese counterparts [Weiss and Kaufman, 2008]. The Bogalusa Heart Study found that a higher BMI and blood pressure in childhood were associated with higher left ventricular mass when they reached adulthood [Lai et al., 2014]. Multiple prospective cohort studies in Finland also asserts this finding [Juonala et al., 2011]. Adults who had always been obese since childhood had an increased risk of type II diabetes and an elevated levels of cardiovascular (CVD) risk biomarkers (i.e. LDL, HDL, triglyceride, and intima-media thickness of carotid artery) compared with adults who always had a normal BMI. Moreover, non-obese adults who were overweight or obese during childhood had similar risk to those who had normal BMI from childhood to adulthood [Juonala et al., 2011]. Even though it is hard to reverse childhood obesity (80% of obese children are reported to become obese adults [Cali and Caprio, 2008]), successfully reversing obesity in children will reduce the risk of developing type II diabetes and CVD later in life. This further put an emphasis on the importance of preventing obesity earlier in life.

2.2 Defining Overweight and Obesity

There are various criteria used in defining childhood overweight and obesity. The Centers for Disease Control and Prevention (CDC) growth chart uses 85^{th} percentile and 95^{th} percentile to define overweight and obesity respectively [Centers for Disease Control and Prevention, 2013]. However, the cut-offs used in the CDC and World Health Organization (WHO) growth chart are designed to represent the US population whom always have had a higher average BMI compared with other countries around the world [Pedrosa et al., 2011, Cole and Lobstein, 2012]. Whereas, the World Obesity Federation (formerly known as the International Obesity Task Force (IOTF)) designs the cut-off points using survey data on children aged 6–18 years from multiple countries (Brazil, United Kingdom, Hong Kong, Singapore, Netherlands, and USA), which are more representative of the world population compared to the CDC. The overweight criteria according to the IOTF cut-off is close to 90^{th} percentile while the obesity cut-off is above the 98^{th} centile [Cole and Lobstein, 2012].

Various anthropometric measurements to define overweight obesity such as BMI, WC, and WHR have a lot of limitations. Even though they have variable strength in predicting adverse health events, a meta-analysis confirmed that no measurement is superior to the others [Vazquez et al., 2007]. It is clear that there is a continuous positive association of increasing weight with the risk of adverse health outcomes. Nevertheless, there is no BMI thresholds that could divide the population into meaningful risk groups [Stommel and Schoenborn, 2010]. In general, the relationship between BMI and all-cause mortality follows a U-shaped pattern. The risk progressively increases as it approach towards both ends of the spectrum. However, the relationship turns into a J-shape among healthy individuals (i.e. the risk of adverse health outcomes in the low BMI group is higher than those in the middle BMI group, but not as high as those in the high BMI group). It was later found that people in the lower BMI spectrum were confounded with their smoking habit, which explains why people in the lower BMI spectrum have a worse outcome [Antonopoulos et al., 2016]. Age is also known confound the relationship between BMI and mortality risk. Being overweight in old age is associated with lower mortality risk compared with those in the normal BMI category [Flegal et al., 2013]. Furthermore, compared with normal-weight group, mortality rate in the overweight and mildly obese group with stable coronary heart disease is lower even after adjusting for confounding factors [Romero-Corral et al., 2006]. Using BMI as a sole indicator in predicting adverse health outcomes without controlling for other factors such as age and health-related behaviours can produce erroneous predictions.

A lot of problems also arise when using universal BMI cut-offs to define overweight/obesity in diverse ethnic groups because body weight distribution and predisposition to store visceral fat tissues differ across ethnic groups. For example, African-American and Hispanic people tend to have the highest average BMI and Asian people have the lowest [Stommel and Schoenborn, 2010, El-Sayed et al., 2011]. In terms of body fat percentage, Chinese, Malaysians, Indians, Indonesian and Japanese people have a higher body fat percentage at a low BMI [Barba et al., 2004]. Also, Māori, Pacific and Indian people had a higher body fat composition compared with the Europeans at the same BMI level. A BMI of 30 kg/m^2 in European men (women) would have an equal body fat percentage as a BMI of 31 (33), 34 (35), and 24 (26) in Māori, Pacific, and Asian Indian respectively [Rush et al., 2009]. Higher body fat percentage has been linked to a higher risk of developing type II diabetes and cardiovascular disease in the Asian population compared with European people of the same age, sex, and BMI [Barba et al., 2004]. A large cross-sectional study in the US also found that the prevalence of chronic health conditions such as hypertension, diabetes, coronary heart disease, asthma, and arthritis varies across ethnic groups regardless of BMI levels. Having said that, after adjusting for socio-demographic and behavioural factors, there are no clear BMI thresholds that divide the population into meaningful risk groups. The only consistent finding is that increasing BMI levels are associated with higher rates of chronic diseases across all ethnic groups [Stommel and Schoenborn, 2010]. It may be better to use ethnic-specific BMI standard deviation when assessing the relationship between BMI and obesity-associated health risk among diverse ethnic groups.

2.3 Lifestyle Interventions

The discourse regarding the cause of the obesity epidemic has been framed on bad dietary habits and low levels of physical activity, suggesting that the responsibility lies on the individuals and can be modified through education as well as physical exercise [Luke and Cooper, 2013]. The Look AHEAD trial, an 8-year multi-centered randomised trial, compared an intensive lifestyle intervention with a less-intensive diabetes support and education on weight reduction. The intervention group achieved maximum weight loss (8.5% of initial weight) in the first year, only to regain their weight back at 4.0%–4.7% loss for the remainder of the trial. Conversely, the comparison group managed to achieve 2.1% weight loss at the end of study [Wadden et al., 2014]. The small differences of weight-loss achieved between intensive lifestyle intervention and usual diabetes support was not substantive enough to make a difference. On top of that, a meta-analysis [Shaw et al., 2006] and long-term clinical trials [de Vos et al., 2016, Wadden et al., 2014] assessing the effect of exercise on overweight and obesity found that most people lost weight initially but regain their weight loss over time, further casting doubt on the long-term effect of exercise and diet-related intervention on maintaining weight loss.

Another problem with lifestyle interventions is their poor compliance rate [Kovacs et al., 2014]. Despite the majority of overweight and obese people wanting to reverse their conditions, expecting them to exercise more also appears to be unrealistic. The National Health and Nutrition Examination Survey (NHANES) measured the duration and intensity of physical activity using an accelerometer, and it found that only 0.3% of adults achieved the current physical activity guideline which recommends 150 minutes of moderate activity or 75 minutes of vigorous activity per week [Luke et al., 2011]. Moreover, some people with a limited functional capacity due to medical conditions are more prone to becoming obese and unable to meet the required physical activity level [Banjare and Bhalerao, 2016]. Considering only a small proportion of the population met the physical activity guideline and the possibility that overweight/obese people would have more difficulty in achieving it, it is very unlikely that increasing population physical activity level would have an impact on obesity prevalence [Luke and Cooper, 2013].

Lifestyle interventions, however, provide health benefits regardless of weight loss [Shaw et al., 2006, Luke and Cooper, 2013]. The results of the Look AHEAD study at four years found that those who were in the intensive lifestyle intervention group achieved greater improvements in blood pressure, triglyceride, and HDL level [Wing et al., 2010]. Furthermore, a 20-year study in China among adults with impaired glucose tolerance found that those who received lifestyle intervention for 6 years had 43% lower diabetes incidence than the control group. Albeit, at the end of 20-year mark, 80% of people in the intervention group had diabetes and the hazard risk ratio of cardiovascular events was inconclusive between the two groups [Li et al., 2008]. In general, a combination of physical exercise and diet is better than either one of them alone in improving markers of cardiovascular disease and diabetes regardless of weight change [Shaw et al., 2006]. Therefore, there is no benefit in questioning the public health value of maintaining a high level of physical activity and healthy diets, and they should be addressed in conjunction with other public health strategies targeting different aspects of health.

Another option for people wanting to reduce their weight is through the pharmacological approach. A meta-analysis on commonly used anti-obesity drugs (Orlistat, Sibutramine and Rimonabant) had been proven to be moderately effective in achieving weight loss. On average, those who used the drugs achieved 5 kg or less of placebo-subtracted weight reductions. Nonetheless, the major barrier to anti-obesity drugs was the lack of adherence, with an average of 30% to 40% attrition rate. The low adherence rate can be explained by potential side effects associated with the use of anti-obesity drugs such as gastrointestinal symptoms, increased blood pressure, and mood disorders [Padwal et al., 2003].

The more invasive strategy for weight loss is through bariatric surgery. In terms of efficacy, the mean ten-year weight reduction ranged from 13% to 25% [Neff et al., 2013]; also, 77% of diabetes, 86% of obstructive sleep apnea, and 62% of hypertension were resolved following the procedure [Buchwald et al., 2004]. A prospective controlled study in Sweden found that the bariatric surgery group had an adjusted hazard risk of 0.67 for cardiovascular events compared with the obese adults receiving usual care (i.e. ranging from lifestyle interventions at many sites to no intervention in other sites) [Sjöström et al., 2012]. Despite its efficacy, bariatric surgery is related to surgical complications such as bowel perforation, bowel obstruction, hemorrhage, anastomotic rupture, micro-nutrient deficiencies, and many others [Neff et al., 2013]. Combined with the fact that only those with a BMI of over 35 kg/m^2 are eligible for this procedure [Jull et al., 2011, Jensen et al., 2013, Sheperd, 2005, Neff et al., 2013] and an operative mortality rate of 0.1–1.1%, making it unlikely to resolve the obesity problem at a population level.

2.4 The Genetics of Obesity

Studies of monozygotic and dizygotic twins have shown that body weight is highly heritable, which means a proportion of the inter-individual difference in body weight can be explained by genetic influence [O'Rahilly and Farooqi, 2008, Herrera and Lindgren, 2010, Barness et al., 2007]. The coefficient of BMI heritability (BMI-H) is found to vary from 31% to 90% depending on the characteristics of the population in the study [Min et al., 2013]. For example, BMI-H is influenced by the average GDP and average BMI of the population. A population with higher average BMI and higher GDP is more likely to have a greater BMI-H. Additionally, BMI-H is also influenced by age. Unique environments are relatively more diverse in adulthood compared with childhood; thus, BMI-H is lower in older population [Min et al., 2013]. This variation happens because the heritability equation assumes that genetic and environmental factors are independent of each other. In reality, we cannot untangle the complex interaction of the gene-environmental relationship. For example, consider a scenario about the behaviour of parents and their children. Genetics may influence the parents' behaviour which in turn influence the environment of the child (e.g. intelligent parents raise their children in an intellectual environment) or the behaviour of the child, which is influenced by genetics, affect their parent's response which is part of the child's environment (e.g. parents respond with frustration to an irritable child) [Stenberg, 2013].

GWAS is designed to detect genetic variants in the genomic loci that are associated with observable characteristics of the population. For example, it can detect associations between common SNPs and common diseases such as heart disease and diabetes [Visscher et al., 2012]. This breakthrough genetic analysis has allowed researchers to capture about 80% of all common variants using as few as 500,000 chosen SNPs. This translates to 20 loci being identified as having an association with body weight [Herrera and Lindgren, 2010]. Among those loci, FTO has been repeatedly identified. Some genes in the FTO region are highly expressed in the brain and play a role in appetite regulations [Mei et al., 2012, Herrera and Lindgren, 2010]. Similar to FTO, other single gene mutations are also shown to mediate the development of obesity through appetite regulating CNS mechanisms [Herrera and Lindgren, 2010, O'Rahilly and Farooqi, 2008], while the rest are known to be involved in fat cell apoptosis and proliferation [Mei et al., 2012].

Despite the success of the GWAS strategy, the established loci only explains around 2% of inter-individual BMI variation [Herrera and Lindgren, 2010, Mei et al., 2012, Bogardus, 2009, Min et al., 2013]. There is a discrepancy in the heritability estimates from the twin studies (31–90%) [Min et al., 2013] and the GWAS approach (around 2%) [Bogardus, 2009], which is called the missing heritability. Although GWASs are agnostic with respect to prior beliefs, they are biased in terms of what is detectable. GWASs are only able to detect causal variants that are common in the population and assume that common diseases are likely to be caused by common variants. However, we cannot dismiss the possibility that rare variants might have a large effect on the population characteristics and they would have been missed by the GWAS approach [Visscher et al., 2012]. Even though GWAS is able to explain the pathophysiology of diseases, the practical utility is questionable. In particular, identifying all 18 types of II diabetes susceptible genes are no better in predicting diabetes than relying only on the person's BMI, age, and sex [Bogardus, 2009]. Further casting doubt on its practicability in predicting disease risk when easier to assess information can make the same prediction.

2.5 Agriculture

The changes in agricultural practice over the past 50 years has provided our world with a greater diversity of foods and less seasonal dependence, thereby changing our dietary practice tremendously. There are two stages of change in the pattern of food consumption, the first one is the "expansion stage", which is followed by the "substitution stage". The expansion stage is characterised by the effort to increase total caloric intake through cheaper foodstuffs rich in carbohydrate. Whereas, in the substitution stage, the total calories consumed do not differ much but there are shifts from carbohydrate-rich staples to vegetable oils and animal products [Kearney, 2010]. These shifts can be seen from the pattern of world food consumption in the past decades. In 1969/71, people around the world on average consumed 2370 kcal/day and the number increased to 2770 kcal/day in 2005/07. The increase was more prominent in low-middle-income countries, despite some parts within the countries still experiencing a food shortage and a high incidence of undernourishment due to unequal food distribution [Alexandratos and Bruinsma, 2012]. Vandevijvere et al. (2015) assessed the relationship between national average food energy supply and body weight using survey data from 56 countries. They found that an increase in food energy supply is significantly associated with an increase in average population weight among high-income countries but not for low- and middle-income country groups. Nevertheless, as food becomes more abundant and a potential decrease in physical activity levels in low- and middle-income countries, the same association between energy intake and body weight is expected to occur in the future [Vandevijvere et al., 2015].

2.6 Urbanisation

Urbanisation may also play a role in shaping the environment that promote the development of obesity. Urbanisation changes how people get their food, where lax trade policies that draw multinational food companies to invest increase the availability of food outlets. As food outlets grow closer to homes, cheap and unhealthy foods become more readily available [Kearney, 2010]. More people in one area also mean that there is a need for a better transportation system, increasing the availability of roads and highways. This, in turn, increases the use of private motorised transportation which reduces physical activities [Malik et al., 2013b, Sturm and An, 2014, Kearney, 2010]. Additionally, technological advancement reduces the need of manual labours as these tasks become more commonly performed by machines and computers. Leisure time is also spent on more sedentary activities such as television viewing, Internet use, and computer gaming [Kearney, 2010, Malik et al., 2013b, Dugas et al., 2011].

In the US, according to a meta-analysis, rural children are 26% more likely to be obese compared with urban children, despite obese rural children being more physically active [Johnson and Johnson, 2015]. Similar findings are also observed in China. Contrary to the stereotype of a higher physical activity level in rural areas, people in rural China tend to have a lower physical activity level due to agricultural mechanisation [Tian et al., 2014]. It was though that rural dwellers have poorer diet due to lack of knowledge and lower level of education. However, a recent study found that urban dwellers consume more animal products rich in fat while having the same total caloric intake than their rural counterparts [Zhang et al., 2017]. This suggests that diets with higher total caloric intake may have a stronger influence than the protective effect of being physically active on obesity.

In contrast, studies in India [Ebrahim et al., 2010] and South-East Asian countries [Angkurawaranon et al., 2014] found that obesity is more prevalent in urban areas. Not only rural men in India had lower obesity prevalence; but they also had lower blood pressure, lipids, and fasting glucose than urban men [Ebrahim et al., 2010]. It seems that the strength of this association is also influenced by the gross national income (GNI) per capita of the countries. Countries with lower GNI per capita, like Vietnam and Laos, had a stronger association; people who lived in urban areas were three times more likely to be obese than those who lived in rural areas. Whereas, rural dwellers were 29% more likely to be obese compared with urban dwellers in Malaysia and Philippines (i.e. countries with higher GNI per capita) [Angkurawaranon et al., 2014]. Surprisingly, despite New Zealand being a high-income country, the association between urban/rural areas and obesity is similar to countries with lower income. In 2002, according to the Children's Nutrition Survey in New Zealand, urban boys were 1.3 times more likely to be obese than those who lived in rural areas, and girls were 1.4 times more likely. There was also no difference in total energy intake per day and the frequency of bouts of physical activity between the two groups [Hodgkin et al., 2010]. These show that there may be underlying mechanisms, other than total energy intake and physical activity, which contribute to the disparity in obesity prevalence between rural and urban areas.

2.7 Food Environment

There have been many studies assessing the role of the environment in shaping the obesity epidemic in recent years. Many environmental studies have looked at how the built environment might affect the levels of physical activity and dietary habits. The current living environment is thought to increase the availability of unhealthy food products (e.g. high-fat foods and sugar-sweetened beverages) and remove spaces for physical activities (e.g. increase private motorised transports) [Giskes et al., 2011]. Despite the sound arguments of environmental theories in trying to explain the obesity epidemic, most of the evidence from epidemiological research has been contradictory [Giskes et al., 2011, Durand et al., 2011, Mattes and Foster, 2014, Sturm and An, 2014].

Studies on accessibility factors (e.g. locations of supermarkets, food store density, availability of healthy foods, etc.) have received considerable attention lately. It was presumed that higher food store density will affect a person's body weight and increase the obesity prevalence in the area [Giskes et al., 2011]. A cross-sectional study in Canada, examining children's weight status and their neighbourhoods' characteristics, found that children had a lower risk of being overweight or obese when they had greater access to more affordable healthy food options within walking distance from their homes [Le et al., 2016]. However, a longitudinal study in the US found no significant impact of food outlet density on children's body weight [Chaparro et al., 2014]. Overall, the relationship between environmental factors and dietary intake have shown that accessibility to supermarkets and the availability of fruits and vegetables in the shops were not associated with fruit and vegetable consumption.

The conflicting evidence emerged due to difficulties and challenges in designing environmental studies. There is no standardised method of assessing the environmental features such as distance to stores and what type of foods were sold in the shops [Giskes et al., 2011]. In addition, assessment on accessibility to certain food shops relies heavily on the assumption that people would buy foods from nearby stores [Le et al., 2016]; discounting the idea that people might travel further to buy their foods.

2.8 Food Price and Consumption

Another important element of the food environment is price. Food price is thought to affect consumption which in turn will influence body weight. Two cohort studies in children have found that one standard deviation increase in fruit and vegetable price is associated with a 4% decrease in consumption [Sturm and Datar, 2011] and a 0.09 increase in a child's BMI z score [Morrissey et al., 2014]. A similar finding is also observed in young adults with an own-price elasticity estimate of -0.32, which is inelastic. However, the elasticity estimate became -0.66 among low-income young adults, indicating that fruit and vegetable consumption in this group is more sensitive to price change [Powell et al., 2009]. Even though the effect of price changes are small, they would have more influence on younger population from a lower income groups.

The association between soft drink/fast food price and food consumption is even less clear. A cohort study in the US found that there is no significant association between fast food or soft drink prices and their consumption among children [Sturm and Datar, 2011]. Moreover, another cohort study assessing the influence of soft drink and fast food prices on children's weight also found that they are not significant predictors for BMI [Morrissey et al., 2014]. The non-significant association between the changes in fast food prices and BMI is also observed among adults. This may have happened because because the price of fast food is already low thus it was unlikely that price changes would have a meaningful influence on consumption. It was also possible that the reduction of fast food consumption due to an increase in price was substituted with an increase in consumption of other food products with the same caloric content, hence no weight change was observed [Cotti and Tefft, 2013].

In general, people perceived the cost of healthier food such as fruit and vegetables to be more expensive and this serves as a major barrier in acquiring them [Dijkstra et al., 2015, Powell et al., 2009, Ni Mhurchu and Ogra, 2007]. A study in the Netherlands has found that lower socio-economic groups adhered less often to healthy dietary guidelines and were more likely to perceived cost as a barrier to buying fruit, vegetables and fish [Dijkstra et al., 2015]. The same situation is also observed in New Zealand. When comparing the average price of regular baskets with the healthier ones showed that it was NZ$36.27–NZ$49.18 more expensive to have a healthier diet. It is not only more expensive, but also finding healthier food options is harder for people living in the less affluent areas. Imposing healthy dietary guidelines to these groups of people will only create frustration and will likely fail in achieving the public health goals [Wang et al., 2010]. Having said that, informing the public about healthy diets such as avoiding saturated fats, reducing total caloric intake, and eating more fruits and vegetables will still be helpful in empowering the public to make better dietary choices. Although, it needs more than just dietary recommendations to halt the obesity epidemic [Woolf and Nestle, 2008].

2.9 Sugar-Sweetened Beverage and Fast Food Consumption on Body Weight

Soft drinks typically refer to beverages that contain carbonated water, but not all soft drinks contain a high level of sugar. A meta-analysis assessing the association between soft drink consumption and body weight found that studies which focused on Sugar-Sweetened Beverage (SSB) had greater effect on weight than those that used the term 'soft drink' [Vartanian et al., 2007]. Higher consumption of SSBs is consistently associated with higher body mass in both adults and children; and the effect is stronger in adults [Malik et al., 2013a, Vartanian et al., 2007]. Apparently, people who consumed SSBs do not compensate for the added energy and tend to increase their total food intake from other sources as well. Not only is sugar known to decrease satiety, but it is also thought to calibrate preference to sweet foods which can lead to an increased intake of sugar-rich food. SSB consumption is also associated with lower consumption of other nutrients such as calcium and vitamins [Vartanian et al., 2007]. On top of that, compared with individuals who consumed none or less than one serving of SSB per month, those who had 1–2 servings of SSB per day had a pooled risk of 1.26 (95%CI: 1.12–1.41) in developing type II diabetes [Malik et al., 2010].

Consumption of fast food, which often contains excessive amounts of refined starched and added sugars, is believed to contribute to obesity. A 15-year prospective study study found that people who consumed more than two servings per week of fast food gained 4.5kg more and had two times greater increase in insulin resistance compared with those who consumed less than one serving per week [Pereira et al., 2005]. However, a study in China did not find any association between fast food consumption and BMI z-score in adolescents [Xue et al., 2016]. It appears that the effect of increased fast food consumption on higher BMI is consistent in adults, but less so in children and adolescents. It is possible that children are still growing and need more energy, thus extra caloric intake is not readily converted into excess body fat [Rosenheck, 2008].

2.10 Sedentary Environment

An environment that promotes sedentary lifestyle was also thought to increase the occurrence of obesity. It was assumed that people living in higher income countries would have lower physical activity and lower energy expenditure as they had more mechanised labour. However, a meta-analysis of doubly-labelled water studies found that energy expenditure did not differ according to the social and economic development of a country despite significant differences in body size, further questioning the lack of physical activity as the primary driver of obesity [Dugas et al., 2011]. Using a model to estimate the change in energy intake that could theoretically contribute to the change in average body size showed that increases in food energy supply alone was sufficient to explain the growth of population weight [Vandevijvere et al., 2015]. A systematic review on built environment also found that the living environment that promotes physical activity (i.e. environment with more mixed land use, open spaces and walk paths) had no significant association with physical activity level or body mass [Durand et al., 2011]. However, most of these studies were cross-sectional and did not take into account neighbourhood self-selection (i.e. physically active people tend to live in an environment that supports their preferences). Even so, the evidence regarding the association between environment and physical activity are mixed even after accounting for neighbourhood self-selection and focusing on quasi experimental designs. Most associations found were in the expected direction or null, which imply that communities with mixed land use or open spaces may increase physical activity of the residents [McCormack and Shiell, 2011]. Despite that, more long-term longitudinal studies are needed to establish a causal link.

2.11 Socio-Economic Status

Research have shown that the socio-economic status is consistently associated with obesity status. Albeit, the direction of the association is dependent on the country's socio-economic level. In low-income countries, the more affluent families had greater overweight or obesity rates. Whereas in high-income countries, they had a lower risk of being overweight or obese [Fruhstorfer et al., 2016, Dinsa et al., 2012, Wu et al., 2015]. This happens because food availability remains a challenge in low-income countries, and only a portion of the communities have more than enough to eat and overweight is still perceived as a sign of wealth [Wu et al., 2015]. In contrast, high-income countries are able to secure enough food even for the lower socio-economic groups, especially those with a strong welfare system. Nevertheless, poor families in these countries do not have the resources to choose healthier diets and lifestyle. Whereas, the wealthier families are able to adhere to healthy diets, which tend to be more expensive [Dinsa et al., 2012, Sturm and An, 2014]. The same pattern of association is also observed in regards to the effect of education level on body weight. Higher education level is associated with a lower risk of obesity in high-income countries, while it is associated with a higher risk of obesity in lower-income countries [Cohen et al., 2013].

A different approach in measuring socio-economic status is by using the degree of the deprivation of an area. Area level deprivation is an aggregated index that measures multiple socio-economic indicators such as income, employment, education, housing, social services, etc. Many studies have shown that area level deprivation is associated with the disparities in obesity risk [Nau et al., 2015, Li et al., 2014, Stafford et al., 2010] and various health outcomes [Maier et al., 2013, Singh et al., 2015, Exeter et al., 2015]. Children who lived in communities with higher socio-economic deprivation have a faster BMI growth compared with those who lived in less deprived areas [Nau et al., 2015]. Children in the deprived communities also had a poor quality diet [Craig et al., 2016], despite the mothers' knowledge about healthy diet being rated as high [Crombie et al., 2008]. The relationship between area level of deprivation and obesity persists even after adjusting for individual and family-socio economic status. This implies that there is a unique pathway in which area level deprivation influences the risk of being obese [Li et al., 2014]. When compared with people who lived in the least deprived areas, those in the most deprived areas also have higher rates of type II diabetes [Evans et al., 2000, Maier et al., 2013].

The Family Food Environment Survey of 136 New Zealand families found that people in the low-income group purchased fewer vegetables despite the fact that all income groups in the survey had basic kitchen amenities and did not have trouble accessing food shops. This shows that the majority of households in New Zealand would have less structural challenges (i.e. lack of facilities to prepare foods) in modifying their dietary habits [Smith et al., 2010]. Nonetheless, according to the NZHS 2014/15, people who lived in the most deprived areas were almost two times more likely to be obese than those who lived in the least deprived areas [Ministry of Health, 2015a]. Furthermore, the result from the NZHS 2015/16 showed that the rates of extreme obesity in people who lived in the least deprived areas have not changed since 2006/07, while people who lived in the most deprived areas experienced an increase in their extreme obesity and obesity rate during the same period [Ministry of Health, 2016].

There are possible explanations to why people who lived in an area with higher deprivation have higher rates of obesity and worse health outcomes. Socio-economically deprived communities may affect health outcomes through regional norms or attitudes towards health [Maier et al., 2013]. Behaviours that are detrimental to health such as smoking, alcohol abuse, and over-consumption of unhealthy energy-dense food may be more prevalent. Lower education level in the area may also result in less emphasis on nutrition and physical activity [Eagle et al., 2012, Nau et al., 2015]. On top of that, deprived communities tend to have higher psychosocial stresses, lower social capital, worse physical infrastructure and an inadequate access to health care services [Maier et al., 2013, Li et al., 2014, Nau et al., 2015, Cummins and Macintyre, 2006]. As opposed to wealthier communities, people in deprived areas would have more challenges in achieving a healthier lifestyle and accessing health care resources, thus they are more at risk of developing obesity and obesity-associated adverse conditions.

2.12 Smoking and Obesity

Multiple studies have found that current smokers tend to be leaner than those who never smoked or had quit smoking [Siahpush et al., 2014, MacKay et al., 2013]. Tobacco contains many harmful components such as nicotine, carbon monoxide, and other stress oxidants. Nicotine binds to the nicotinic cholinergic receptors in the brain and activates the sympathetic nervous system, which then causes an increase in heart rate, blood pressure, and cardiac work. The increase in sympathetic activity also increases energy expenditure which explains the lower body weight observed among current smokers [Rigotti and Clair, 2013]. Age also plays an important role in this relationship. A study in Scotland found that people who had smoked more than 20 years were 50% more likely to be overweight than those who had never smoked. They also found that those aged 16–24 years who quit smoking were no more likely to be overweight or obese than current smokers or people who never smoked [MacKay et al., 2013]. Although current smokers tended to have normal weight, their overall mortality risk was higher than overweight or obese ex-smokers [Siahpush et al., 2014]. Smoking cessation among the younger population is desirable because it is not associated with weight gain and would reduce the mortality risk. Despite potential weight gain associated with smoking cessation among older population, it should not deter them from quitting considering an overall protective effect from mortality that they would gain.

2.13 Healthy Immigrant Effect

Many studies in the US, Canada, Australia, and the United Kingdom have found that immigrants have a better health status than the citizens of the recipient-countries [Vang et al., 2015, Kennedy et al., 2015, Goel et al., 2004]. In terms of obesity, a study in the US found that recent immigrants had significantly lower prevalence of obesity compared with the US-born citizens. However, the obesity rates of the immigrants began to climb up after 10 years and caught up with the host prevalence after 15 years of living in the US [Goel et al., 2004]. The protective effect of being an immigrant is less clear on children and is reversed for expecting mothers, older immigrants [Vang et al., 2015], and refugees [Norredam et al., 2014]. In general, the healthy immigrant effect can be explained by the immigration selection favoring highly educated and healthier immigrants [Kennedy et al., 2015]. Nonetheless, this selection process is less relevant in the perinatal period, childhood, adolescence, elderly, and refugees; which explains why the healthy migrant effect is not observed in these groups [Vang et al., 2015, Norredam et al., 2014].

2.14 Literature Review Summary

This chapter addresses that obesity is an important issue because it is associated with a higher risk of developing chronic conditions such as type II diabetes and CVDs. Defining overweight and obesity in the population is not easy, especially when dealing with diverse ethnic groups as they carry different risks at the same BMI level. It is also clear that tackling obesity at a population level through individual interventions are not enough, therefore, studying which environmental factors that have a strong influence on obesity is necessary to devise better public health interventions. Multiple studies have found that people who lived in more deprived areas had a higher likelihood of becoming overweight/obese. It is important to study this relationship to reduce the health disparities across areas with different area level deprivations. This study will analyse this relationship while controlling for possible confounding factors which I have discussed in this chapter.

3 Population and Methods

3.1 Theoretical Framework

Obesity prevailed despite efforts to reduce obesity rates through lifestyle and pharmacological interventions. Interventions that try to improve physical activity levels and dietary habits fail to sustain long-term weight loss [Padwal et al., 2003, Shaw et al., 2006, Wadden et al., 2014]. Environmental factors play an important role in the obesity epidemic because they determine accessibilities to unhealthy/healthy food, open spaces for physical activity and health care services [Giskes et al., 2011, Sturm and An, 2014]. Areas with high deprivation level have social and structural challenges that put people at a higher risk of being overweight/obese [Nau et al., 2015, Li et al., 2014, Stafford et al., 2010]. The theory which serves as the foundation of this research is shown in Figure 3.1 below.

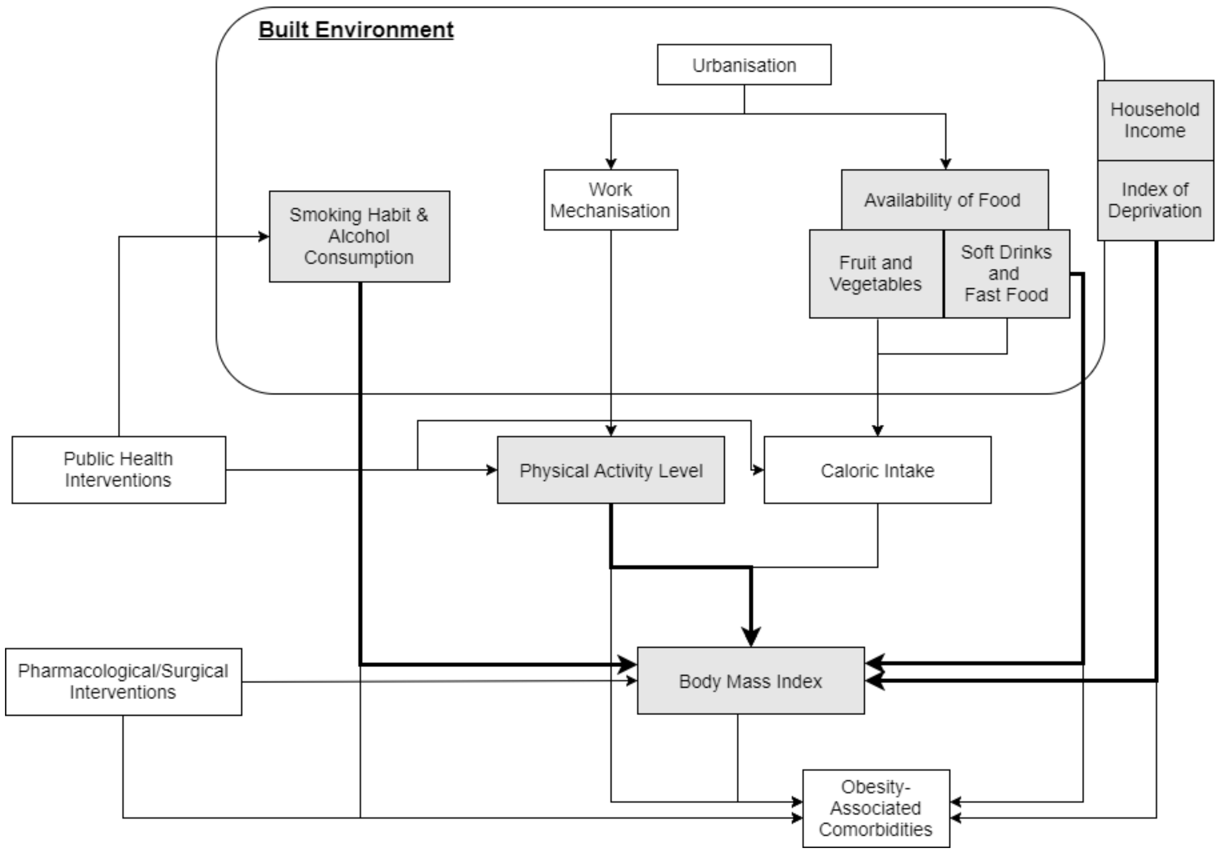

Figure 3.1: Theoretical Framework. Urbanisation is thought to change the built environment, which in turn modifies the behaviours of the people. Work mechanisation and the changes in the availability of certain food may influence physical activity levels and the caloric intake of the population, all of which would impact on the BMI. Also, income and area level deprivation through their relationship with the built environment are thought to have an effect on BMI and obesity-associated comorbidities. Grey boxes indicate available data and bold lines are the main focus of this study.

3.2 Data

I obtained data on age, ethnicity, weight, height, smoking status, drinking problem, physical activity level and deprivation quintile from the confidentialised unit record files (CURFs) of NZHSs. The NZHSs were conducted in 1996/97, 2002/03, 2006/07, 2011/12, 2012/13, 2013/14 and 2014/15. This study excluded the NZHS 1996/97 from the analysis because it did not have the information on the height and weight of individuals. I dropped the data on individuals below 15 years old from the Child Nutrition Survey 2002 (which was intended to supplement the NZHS 2002/03) because it used a different set of questions.

The NZHS is intended to capture the health status of a representative sample of New Zealand residents, which is generalisable to the New Zealand population. It is administered to individuals aged 15 years and above, as well as to children 0 to 14 years, generally through their primary caregiver. The sampling design is multi-stage, stratified, probability-proportional-to-size with meshblocks as the primary sampling units. A meshblock is the smallest geographically defined area unit for which statistical data is reported by Statistics New Zealand. Statistical weighting is used to ensure the data are representative of the target population, including the sub-groups of population. The content of the health surveys has been preserved so that the data can be compared over time [Ministry of Health, 2015b]. After excluding respondents younger than two years old, the total sample size obtained from the NZHSs was 95,034 respondents.

3.3 Hypotheses

Area level deprivation is a variable used to measure the degree of deprivation a certain area has. In New Zealand, it is calculated using the following information: communication access, income support, income level, employment, educational qualification, home ownership, supports, living space, and means of transportation. The New Zealand Deprivation Quintile is a relative measure of area level deprivation which ranges from one (the least deprived) to five (the most deprived) [Atkinson et al., 2014]. Higher deprivation quintile is associated with a higher overweight/obesity prevalence [Ministry of Health, 2015a, Ministry of Health, 2016]. Thus, I formulated my hypotheses as follows:

- **Hypothesis 1**

  Null (H_0): The likelihood of overweight/obesity is the same among residents of areas with different deprivation quintiles.

  Alternative (H_1): The likelihood of overweight/obesity is higher among residents of areas with higher deprivation quintiles.

- **Hypothesis 2**

  This second set of hypotheses is intended to control for confounding factors.

  Null (H_0): The likelihood of overweight/obesity is the same among residents of areas with different deprivation quintiles after adjusting for age, sex, ethnicity, household income, education and health-related behaviours.

  Alternative (H_1): The likelihood of overweight/obesity is higher among residents of areas with higher deprivation quintiles after adjusting for age, sex, ethnicity, household income, education and health-related behaviours.

3.4 Variables

3.4.1 Outcome Variables

- Overweight/Obese

  The parameter used to measure overweight and obesity is the BMI. BMI is a person′ weight in kilograms divided by the square of height in meters. In adults, overweight is defined by a BMI greater than or equal to 25, and obesity is defined by a BMI greater than or equal to 30 [Centers for Disease Control and Prevention, 2013]. The International Obesity Task Force (IOTF) BMI cut-offs were used to define overweight and obesity in participants aged 2–17 years [Ministry of Health, 2015b, Cole et al., 2000, Cole and Lobstein, 2012]. The IOTF BMI cut-offs table can be obtained from the World Obesity Federation (formerly known as IOTF) website [World Obesity Federation,].

3.4.2 Explanatory Variables

- New Zealand Deprivation Quintile

  That is an ordinal scale from one (least deprived) to five (most deprived) which measures nine dimensions of area level deprivations.

3.4.3 Covariates

- Physically Active

  This is a dichotomous variable which tells whether people aged 18 years or above had 30 minutes of brisk walk or moderate physical exercise (15 minutes if it is a vigorous physical activity) on five or more days in the past seven days.

- Sedentary Lifestyle

  This is a dichotomous variable that tells whether people aged 18 years or above had not had at least 30 minutes of exercise in the past seven days.

- Fruit Guideline

  This is a dichotomous variable that tells whether people met the dietary fruit guideline, which is two or more servings of fruit per day.

- Vegetable Guideline

  This is a dichotomous variable that tells whether people met the dietary vegetable guideline, which is three or more servings of vegetable per day.

- Migration Status

  This is a dichotomous variable that tells whether people aged 18 years or more had migrated to New Zealand less than 10–11 years ago at the time of the survey.

- Urban or Rural Area

  Urban area is defined as an area that has a population of at least 1000 people. Rural area is defined as an area that has a population of less than 1000 people.

- Educational Qualification

  People who had either secondary, tertiary, or no educational qualification. In children, it refers to the educational qualification achieved by the parents, whomever the highest.

- Difficulty Climbing Several Flights of Stairs

  People aged 18 years or more who had either a lot of difficulties, a little difficulties or no difficulty climbing several flights of stairs.

- Smoking status

  People aged 18 years or above who were either non-smokers, ex-smokers, or current smokers. Current smokers were people who had smoked at least 100 cigarettes over a life-time and were currently smoking at least monthly. Ex-smokers were people who had smoked more than 100 cigarettes over a life time and had stopped for more than a month from the date of interview. Non-smokers were people who were not elsewhere included.

- Drinking Problem

  People who had an Alcohol Use Disorders Identification Test (AUDIT) score greater than seven. AUDIT is a screening tool developed by the WHO to assess alcohol consumption, drinking behaviors, and alcohol-related problems. A score greater than seven is considered to indicate harmful alcohol use.

- Soft Drink Consumption

  People aged 2–14 years who consumed soft drink in one of the following patterns: none, one, two to three or more than four times per week. Soft drink here includes cola, lemonade or sugar-sweetened energy drinks; but excludes powdered cordial made with water or plain fruit juice. The emphasis is on sugar-sweetened beverages.

- Fast Food Consumption

  People aged 2–14 years who consumed fast food in one of the following patterns: none, one, two to three, or more than four times per week. Fast food here includes fish and chips, pizza, burger, fried chicken or other takeaways.

- Household Income

  People who belonged to one of the four household income categories: $\leq$ \$15,000; \$15,001 to \$40,000; \$40,001 to \$70,000; and $\geq$\$70,001.

- Year

  That is the period when the New Zealand Health Surveys were conducted.

- Age

  It is a numerical scale derived from the date of birth and the date of the interview in years.

- Sex

  People who identified as either male or female.

- Ethnicity

  People who identified themselves as belonging to Māori only, Pacific only, Asian only (Indian and/or Chinese), European only, 2+ ethnicities (M) (belonging to two or more ethnic group with Māori ancestry), or Other (not elsewhere included).

3.5 Data Analysis

3.5.1 Data Cleaning

Overweight and obesity categories in adults were defined using the WHO criteria. According to this, an individual is defined as overweight if the BMI score is between $\geq$25 and 30 kg/m2, and obese if the BMI score is $\geq$30 kg/m2. 1.2kg is added to the weight variable to remove the clothing adjustment made in the NZHS 2002/03. The existing overweight and obesity category in the NZHS 2002/03 used higher cut-offs for Māori and Pacific people. In the later surveys, the WHO overweight/obesity cut-offs were used for everyone regardless of their ethnicity. New code was written to create overweight and obesity category according to the WHO criteria for NZHS 2002/03 so that the data would be comparable to the rest of the health surveys.

The cut-offs used for child overweight and obesity in the NZHS 2006/07, 2011/12, and 2012/13 were based on the old IOTF cut-offs. Even though the old and new IOTF cut-offs do not differ significantly in defining overweight and obesity [Cole and Lobstein, 2012], this study decided to use the new IOTF cut-offs based on the children's age in years across the data sets to be consistent. The new IOTF cut-offs based on the children's age in years were also applied to the NZHS 2013/14 and 2014/15, because they used the cut-offs based on the children's age in months.

Children aged 0–14 years are recorded in the child NZHS, while 15+ years old are recorded in the adult NZHS. I combined and then split the data into the Child data (2–17 years old) and the Adult data (18+ years old).

I did this so that the analysis on adults would be easier as the overweight/obesity criteria does not vary by age starting from 18 years old.

Imputations on missing income data were only done in the NZHS 2002/03, but not in any of the other survey periods. Therefore, I decided not to use any imputation in my analysis. Imputed income data in the NZHS 2002/03 were removed and the missing income data from the rest of the data sets were excluded list-wise from the analysis. Although there was no household income data and parent's educational level in the NZHS 2011/12 and 2012/13 child data, this information could be obtained through matching the household ID (HHID) from the adult data of the same survey period. After matching the HHID, there were two different educational level data for the same child. This discrepancy was not addressed in any of the NZHS methodology reports published by the Ministry of Health and may had happened because there were two parents who had different qualifications. I used the highest parental education level in this case.

The NZHSs used ethnic prioritisation in their ethnicity variable, which means if someone identified themselves as having multiple ethnicities, the ethnicity variable would have been forced into one ethnicity with a higher priority. The order of priority was Māori, Pacific, Asian, and European/other ethnicity. For example, people who identified as belonging to Māori and Asian descents would be treated as being in the Māori only group. Even though ethnic prioritisation can simplify the analysis, it is not ethnically neutral and eliminates the information that can be gained from younger people who more frequently belong to multiple ethnic groups [Didham and Callister, 2012]. Codes for a new ethnic groupings were written, and the new groupings were broken down into Māori only, Pacific only, Asian only, European only, 2+ ethnicities (M), and "Other" group. Other Asians (besides Indian and/or Chinese) were included in the "Other" category because there was no option for the respondents to choose other Asian ethnicities in the NZHSs.

The R codes used to clean the NZHS data are presented in Listing A.1.

3.5.2 Descriptive and Bivariate Analyses

I conducted descriptive statistical analysis of single variables after applying the sample weights. The nominal or ordinal variables were reported as count and percentages while the numeric variables were reported as five numbers summary (mean, minimum, maximum, first inter-quartile and third inter-quartile) in a table. Unweighted bivariate analyses were conducted on all unique combinations between explanatory variables and the outcome variable. Chi-squared test was used when both the explanatory and the outcome variable were categorical. ANOVA test was used when the explanatory variable was numeric and the outcome variable was categorical. All statistical results from the bivariate analyses along with their effect sizes were reported. I also conducted bivariate analyses on each survey period separately and the R codes are presented in Listing A.3 and A.4. In addition, I plotted the prevalence of overweight/obesity over time, and the age-standardised obesity prevalence by ethnicity over time.

3.5.3 Regression Modelling

The analysis in this study used the sample weights so that the results would be representative to the whole population. The NZHS 2006/07 did not have unique identifiers for its cluster variable, thus unique clusters were created for the NZHS 2006/07 as to be consistent with the remaining health surveys. The complex survey design object was created using the 'survey' package in R [Lumley, 2016, Lumley, 2004]. The R codes used to do the ordinal regression analyses are presented in Listing A.2.

A forward stepwise proportional odds ordinal regression using complex survey design was conducted to find the final model. My interest was in the effect of area level deprivation on overweight/obesity outcome. In the first step, deprivation quintile and linear contrast were added into the model. Then, I added one covariate at a time and dropped the covariate if it did not reach a p-value of at least 0.1. Proportional odds assumption was assessed by running a binomial general linear model with contrasts on different threshold values. I followed the instruction to assess proportional odds assumption by 'coreysparks' [coreysparks, 2015] and the R codes were also presented in Listing A.2.

The proportional odds of cumulative logit model is given in the following equation [Agresti, 2013]:

$$logit[P(Y \geq j|x_1, ..., x_i)] = \alpha_j + \beta_1^T x_1 + ... + \beta i^T x_i, \qquad (j = 1, ..., J-1;\ i = 1, ..., n) \tag{3.1}$$

Where:

Y = outcome variable

j = threshold point which divide Y into two categories

x = explanatory variable

α_j = intercept coefficient for j threshold

β_i^T = beta coefficient for each explanatory variable x, assumed to be the same for each logit

$P[Y \geq j|x_1, ..., x_i]$ = probability of Y outcome above j threshold given explanatory variable $x_1, ..., x_i$

4 Results

4.1 Descriptive Analysis of The NZHS

The descriptive summary of the Adult data is presented in table 4.1 after applying the sample weights. The prevalence of obesity had not changed statistically significantly since 2011/12. The prevalence in 2014/15 (32%) was higher than in 2002/03 (26%) ($p < .001$) and in 2006/07 (28%) ($p < .001$) (Figure 4.1). In regards to ethnicity, the proportion of European group decreased while the Other group increased over time. The proportion of respondents who lived in urban areas was higher in 2014/15 (86.8%) than in 2002/03 (80.1%) ($p < 0.01$) and 2006/07 (86%) ($p < .001$). There is no data on the urban/rural information in the 2011/12 and 2013/14 NZHS, thus no conclusion could be made on those years. In regards to health-related behaviours, the proportion of people who met the fruit guideline in 2014/15 (55%) was lower than the figure in 2006/07 (60%) ($p < .001$), and it was not different than in 2002/03 (54%) ($p = .2$). The proportion of people who met the vegetable guideline decreased from 69% in 2011/12 to 66% in 2014/15 ($p < .001$). There was no statistically significant difference in the proportion of people who were physically active across the survey periods except in 2011/12 where it peaked at 56%. The proportion of people who lived a sedentary lifestyle in 2014/15 (14%) was higher compared with the figure in 2011/12 (11%) ($p < .001$), but it was lower than in 2002/03 (16%) ($p < .01$). The BMI in the Adult data had a skewness of 1.26 and a kurtosis of 5.62 ($p < .001$), which means the BMI data was skewed to the right and leptokurtic. Therefore, I categorised the BMI variable in the Adult data into tertiles.

A summary of descriptive statistics for the Child data is presented in Table 4.2. The proportion of obese children in 2012/13 (13%) was higher than in 2006/07 (10%) ($p = .014$). However, compared with the proportion of obese children in 2014/15 (13%), there was no difference since 2011/12 ($p < .001$) (Figure 4.1). Similar to the Adult data, the proportion of the European group also decreased while the proportion of the Other ethnic group increased over time. The proportion of children who lived in urban areas was lower in 2014/15 (80%) compared with the figure in 2006/07 (85%) ($p = .01$). The proportion of parents who had tertiary education was higher in 2014/15 (44%) compared with the figure in 2012/13 (39%) ($p < .01$). There was no significant statistical difference in the proportion of children who had soft drink or fast food more than once per week across time. Missing data in both Adult and Child data are presented in Table 4.3.

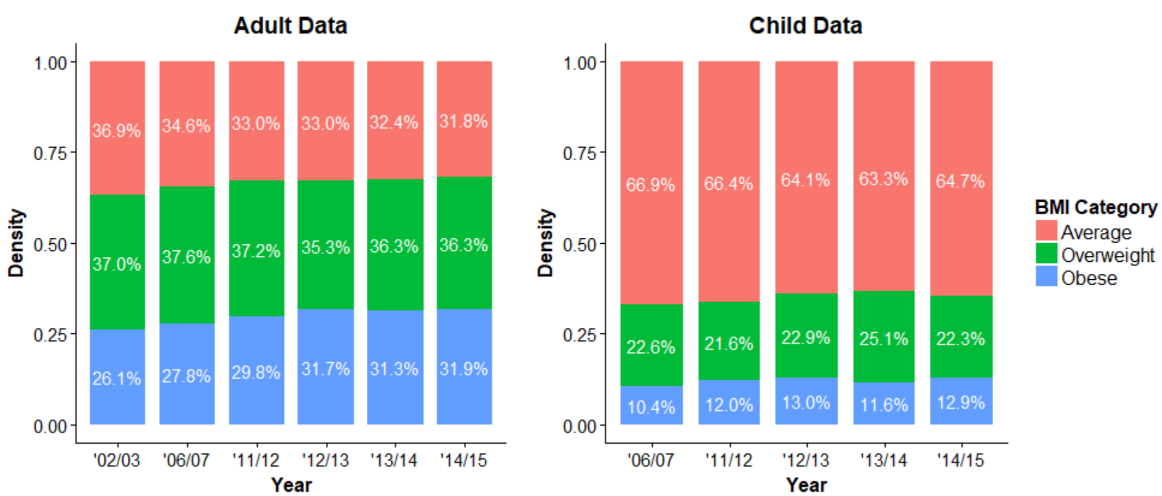

Figure 4.1: BMI Categories in The Adult and Child Data

Table 4.1: Descriptive Statistics of The Adult Data (2002/03–2014/15 NZHS)

Variables		New Zealand Health Survey Period					
		2002/03	2006/08	2011/12	2012/13	2013/14	2014/15
Sample Size		10812	11010	10234	11444	12046	12185
Age (in Years)	**Minimum**	18	18	18	18	18	18
	1st Quartile	31	33	32	32	32	32
	Median	43	44	46	46	46	46
	3rd Quartile	56	58	59	60	61	61
	Max	97	95*	90*	90*	90*	90*
BMI (kg/m^2)	**Minimum**	18.53	18.51	18.50	18.50	18.50	18.50
	1st Quartile	23.59	23.81	24.06	23.99	24.12	24.16
	Median	26.54	26.86	26.98	27.18	27.31	27.32
	3rd Quartile	30.20	30.53	30.91	31.38	31.21	31.27
	Max	64.21	80.56	77.01	68.20	70.86	78.02
		% (SD)					
BMI Category	**Average**	36.88 (0.70)	34.60 (0.64)	32.98 (0.74)	32.96 (0.69)	32.40 (0.66)	31.80 (0.69)
	Overweight	37.02 (0.65)	37.58 (0.59)	37.21 (0.69)	35.32 (0.66)	36.31 (0.62)	36.34 (0.62)
	Obese	26.09 (0.64)	27.82 (0.58)	29.80 (0.69)	31.72 (0.65)	31.30 (0.68)	31.86 (0.68)
Sex	**Male**	49.99 (0.65)	49.72 (0.7)	49.84 (0.64)	50.87 (0.69)	49.74 (0.63)	49.34 (0.62)
Ethnic Group	**European Only**	78.72 (0.87)	73.93 (0.67)	60.30 (0.90)	62.43 (0.85)	62.41 (0.85)	60.86 (0.85)
	2+ Ethnicities (M)	3.93 (0.27)	5.30 (0.22)	5.38 (0.28)	5.02 (0.26)	5.17 (0.25)	5.54 (0.30)
	Asian Only	3.82 (0.33)	5.83 (0.35)	5.46 (0.49)	5.80 (0.42)	6.27 (0.46)	6.95 (0.49)
	Maori Only	6.31 (0.38)	5.58 (0.26)	6.24 (0.34)	7.27 (0.36)	6.97 (0.36)	6.69 (0.33)
	Pacific Only	3.96 (0.45)	4.12 (0.34)	3.21 (0.30)	3.62 (0.32)	3.37 (0.30)	3.63 (0.30)
	Other	3.26 (0.27)	5.24 (0.28)	19.41 (0.66)	15.86 (0.55)	15.79 (0.54)	16.33 (0.53)
Deprivation Quintile	**1**	20.45 (1.88)	18.4 (1.31)	21.73 (1.34)	20.72 (1.84)	20.78 (1.76)	20.24 (1.83)
	3	20.91 (1.64)	20.88 (1.5)	20.56 (1.2)	20.63 (1.68)	20.32 (1.6)	20.75 (1.5)
	5	18.45 (1.19)	19.39 (1.5)	17.62 (1.08)	18.04 (1.3)	18.24 (1.26)	18.29 (1.22)
Urban/Rural Area	**Urban**	80.13 (1.45)	86.01 (1.25)	-	-	-	86.75 (0.96)
Household Income	**<= $15k**	10.1 (0.49)	5.87 (0.28)	3.55 (0.3)	4.16 (0.36)	3.78 (0.28)	3.69 (0.26)
	$15–40k	32.75 (0.85)	25.15 (0.56)	21.19 (0.7)	23.07 (0.69)	21.04 (0.67)	21.88 (0.67)
	$40–70k	27.03 (0.69)	25.52 (0.56)	25.11 (0.7)	24.97 (0.67)	24.19 (0.68)	24.27 (0.67)
	>$70k	30.12 (0.91)	43.46 (0.78)	50.15 (1.03)	47.81 (0.94)	50.99 (0.92)	50.16 (0.97)
Educational Qualification	**None**	20.73 (0.65)	16.88 (0.47)	14.81 (0.52)	15.64 (0.51)	14.06 (0.48)	13.89 (0.46)
	Secondary	28.36 (0.64)	26.41 (0.58)	31.68 (0.68)	32.45 (0.66)	34.45 (0.65)	33.49 (0.67)
	Tertiary	50.91 (0.8)	56.71 (0.67)	53.51 (0.83)	51.91 (0.76)	51.49 (0.75)	52.62 (0.77)
Meeting Fruit Guideline	**Yes**	54.1 (0.75)	59.6 (0.63)	58.44 (0.72)	58.09 (0.69)	56.42 (0.67)	55.37 (0.67)
Meeting Vegetable Guideline	**Yes**	69.14 (0.83)	65.04 (0.68)	69.23 (0.79)	67.44 (0.8)	64.91 (0.76)	65.52 (0.74)
Migration Status	**Migrant**	4.46 (0.31)	8.67 (0.39)	11.13 (0.56)	10.26 (0.5)	10.63 (0.53)	11.01 (0.53)
Difficulty Climbing Several Flights of Stairs	**A lot**	8.69 (0.4)	9.64 (0.33)	5.66 (0.3)	6.46 (0.28)	7.04 (0.32)	7.02 (0.32)
	A little	13.5 (0.48)	16.82 (0.45)	10.4 (0.4)	10.95 (0.4)	10.06 (0.38)	11.19 (0.43)
	None	77.81 (0.48)	73.53 (0.45)	83.94 (0.4)	82.59 (0.4)	82.89 (0.38)	81.79 (0.43)
Smoking Status	**Non Smoker**	-	55.5 (0.7)	54.86 (0.73)	55.13 (0.69)	55.11 (0.67)	55.74 (0.66)
	Ex Smoker	-	24.14 (0.54)	26.27 (0.55)	26.8 (0.57)	26.97 (0.56)	27.23 (0.58)
	Current Smoker	-	20.36 (0.54)	18.87 (0.58)	18.07 (0.54)	17.91 (0.53)	17.03 (0.52)
Drinking Problem	**AUDIT >7**	17.5 (0.63)	18.13 (0.54)	15.74 (0.58)	16.06 (0.55)	16.81 (0.57)	18.53 (0.61)
Physically Active	**Yes**	53.12 (0.81)	52.85 (0.68)	56.18 (0.96)	52.77 (1.01)	53.03 (1)	51.55 (0.82)
Sedentary Lifestyle	**Yes**	16.21 (0.61)	13.22 (0.4)	10.86 (0.5)	12.98 (0.53)	13.36 (0.52)	13.75 (0.47)

* Maximum age is capped at this number

Table 4.2: Descriptive Statistics of The Child Data (2006/07–2014/15 NZHS)

Variables		New Zealand Health Survey Period				
		2006/08	2011/12	2012/13	2013/14	2014/15
Sample Size		4415	3310	3415	3788	4020
Age (in Years)	**Minimum**	2	2	2	2	2
	1st Quartile	6	6	6	6	6
	Median	10	10	10	10	10
	3rd Quartile	14	14	14	14	14
	Maximum	17	17	17	17	17
		% (SD)				
Sex	**Male**	51.61 (1.01)	52.49 (1.26)	50.68 (1.16)	51.95 (1.17)	50.32 (1.11)
BMI Category	**Average**	66.94 (1)	66.39 (1.27)	64.1 (1.19)	63.35 (1.08)	64.73 (1.18)
	Overweight	22.65 (0.86)	21.6 (0.96)	22.95 (1.03)	25.07 (0.92)	22.33 (0.94)
	Obese	10.41 (0.61)	12.01 (0.79)	12.96 (0.84)	11.58 (0.7)	12.94 (0.72)
Ethnic Groups	**European Only**	60.33 (1.14)	48.08 (1.57)	51.02 (1.53)	46.81 (1.42)	48.93 (1.46)
	2+ Ethnicities (M)	13.29 (0.58)	12.35 (0.73)	13.55 (0.78)	15.05 (0.78)	14.49 (0.76)
	Asian Only	4.01 (0.37)	4.53 (0.63)	4.5 (0.57)	5.11 (0.6)	4.48 (0.51)
	Maori Only	8.77 (0.5)	10.29 (0.84)	11.03 (0.77)	10.53 (0.71)	11.49 (0.75)
	Pacific Only	7.19 (0.68)	7.57 (0.92)	6.81 (0.76)	5.9 (0.65)	5.65 (0.57)
	Other	6.42 (0.68)	17.18 (0.92)	13.07 (0.76)	16.59 (0.65)	14.94 (0.57)
Deprivation Quintiles	**1**	21.87 (1.51)	22.11 (2.16)	20.59 (1.97)	20.95 (2.13)	19.83 (2.14)
	3	19.38 (1.22)	16.7 (1.55)	18.74 (1.63)	17.13 (1.52)	17.52 (1.55)
	5	20.49 (1.27)	23.41 (1.82)	23.29 (1.62)	23.68 (1.61)	23.69 (1.58)
Urban/Rural	**Urban**	85.06 (1.15)	-	-	-	80.06 (1.67)
Household Income	**<= $15k**	4.68 (0.42)	3.78 (0.63)	6.2 (1.33)	4.06 (0.56)	3.21 (0.45)
	$15–40k	23.15 (0.94)	16.92 (1.32)	17.85 (1.18)	17.83 (1.03)	18.04 (1.03)
	$40–70k	29.91 (1.05)	23.71 (1.43)	25.64 (1.28)	26.35 (1.32)	23.06 (1.07)
	>$70k	42.26 (1.23)	55.59 (1.86)	50.31 (1.76)	51.76 (1.57)	55.69 (1.48)
Parents Qualification*	**None**	19.34 (0.86)	19.54 (1.22)	21.3 (1.13)	14.43 (0.92)	15.64 (0.88)
	Secondary	35.74 (1.07)	40.12 (1.41)	40 (1.27)	42.31 (1.27)	40.71 (1.19)
	Tertiary	44.93 (1.09)	40.35 (1.46)	38.7 (1.37)	43.26 (1.2)	43.65 (1.24)
Meeting Dietary Guideline	**Fruit**	61.44 (2.79)**	68.48 (1.2)	71.01 (1.16)	69.38 (1.09)	70.87 (1.04)
	Vegetable	53.01 (2.76)**	60.05 (1.33)	58.41 (1.28)	57.44 (1.27)	58.04 (1.19)
Soft Drink Consumption (per week)*	**0**	36.47 (1.08)	36.78 (1.29)	36.9 (1.16)	40.47 (1.25)	39.74 (1.25)
	1	28.03 (0.94)	28.48 (1.12)	28.39 (1.12)	27.68 (1.07)	27.65 (1.03)
	2–3	23.62 (0.89)	21.44 (1.07)	23.01 (1.14)	21.36 (0.93)	20.37 (1.02)
	4+	11.89 (0.68)	13.3 (0.9)	11.71 (0.69)	10.5 (0.68)	12.25 (0.75)
Fast Food Consumption (per week)*	**0**	28.92 (0.98)	31.49 (1.12)	33.11 (1.12)	32.36 (1.25)	30.45 (1.07)
	1	50.75 (1.03)	48.22 (1.23)	47.59 (1.22)	47.85 (1.26)	49.19 (1.13)
	2–3	16.95 (0.78)	17.82 (0.95)	17.08 (0.85)	17.79 (0.88)	18.1 (0.89)
	4+	3.38 (0.4)	2.47 (0.4)	2.23 (0.31)	2 (0.31)	2.27 (0.29)

* Below 15 years old
** Data on children younger than 15 years old is not available

Table 4.3: Missing Data Table

Missing Data*	2002/03	2006/07	2011/12	2012/13	2013/14	2014/15
			N (%)			
Adult BMI	1189 (9.78)	788 (6.68)	1864 (15.41)	1038 (8.32)	696 (5.46)	713 (5.53)
Child BMI	-	177 (3.86)	668 (16.77)	543 (13.72)	372 (8.94)	301 (6.97)
Adult Household Income	2807 (23.09)	1393 (11.81)	3920 (32.40)	3651 (29.25)	3639 (28.56)	2220 (17.21)
Child Household Income	-	614 (13.37)	1612 (40.47)	1445 (36.51)	1379 (33.15)	958 (22.17)
Parent's Education	-	-	194 (4.87)	246 (6.22)	195 (4.69)	167 (3.87)

* Only showing variables with 3% or more missing data

4.2 Bivariate Analyses

I did not use sample weights in these analyses. Figure 4.2 and Figure 4.3 show the bivariate charts on all possible combinations between the covariates and the BMI outcomes, which are BMI tertiles in adults and BMI categories in children. The bivariate tables are presented in Appendix B. Summary of the bivariate statistical tests are presented in Appendix B.3. Even though most explanatory variables were statistically significantly associated with the outcome variables, only several variables had a small or medium effect size according to the Cohen's guide [Cohen, 1992]. Age (η^2= 0.01), sex (φ_c= 0.11) and difficulty climbing stairs (φ_c= 0.11) had a small size effect in the Adult data, while deprivation quintile and education had a small effect size in both the Adult (φ_c= 0.12; φ_c= 0.09) and Child data (φ_c= 0.11; φ_c= 0.07). Ethnicity variable was the only variable that had a medium effect size in the Adult (φ_c= 0.22) and Child data (φ_c= 0.15). Separate bivariate analyses on each survey period are presented as R codes in Listing A.3 and A.4.

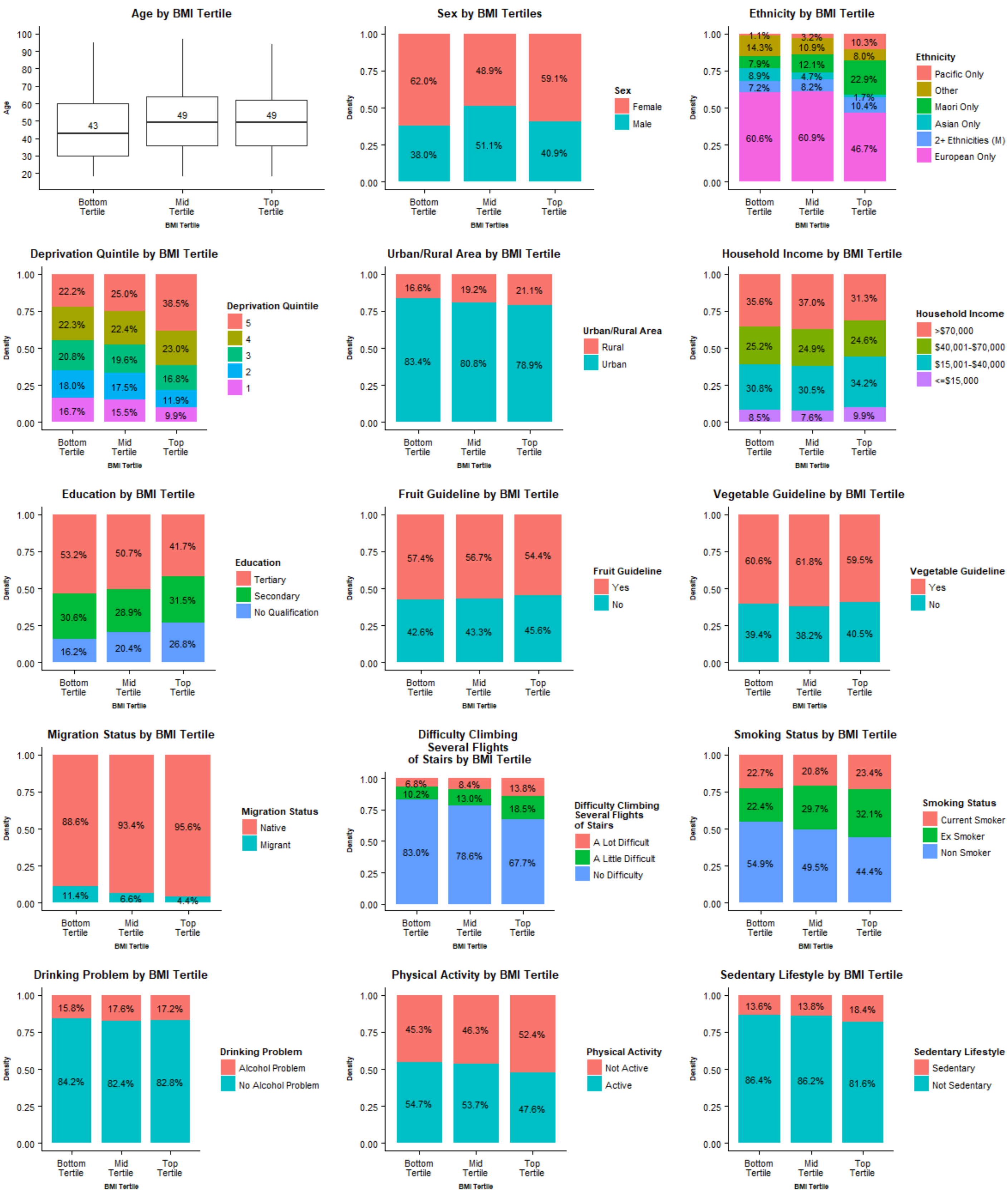

Figure 4.2: Bivariate Charts for The Adult Data

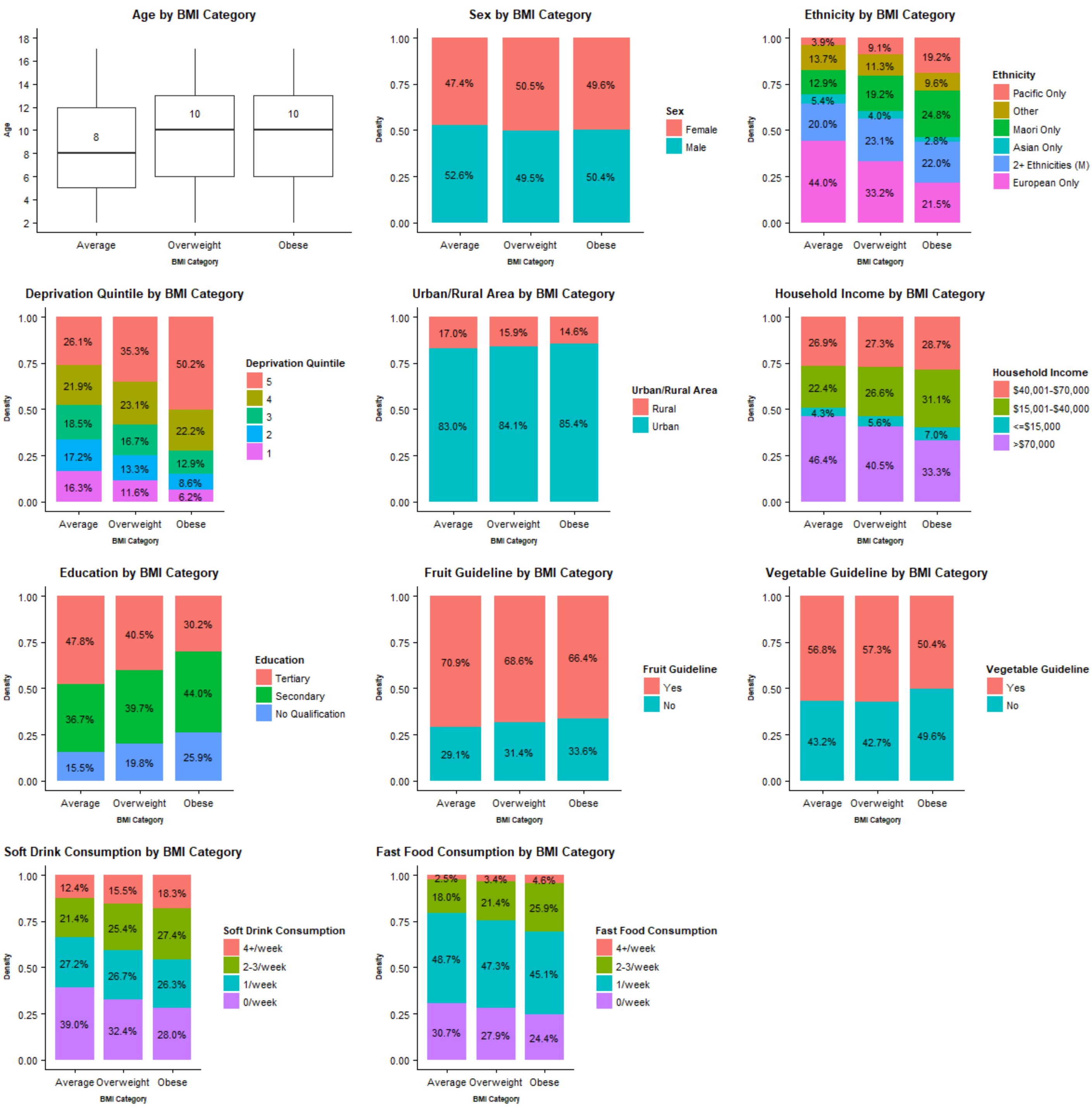

Figure 4.3: Bivariate Charts for The Child Data

4.3 Obesity Prevalence across Ethnic Groups

Figure 4.4 shows ethnic-specific obesity prevalence over time in adults and children. In adults, almost three out of every four Pacific adults were obese in 2014/15 and this was more than twice the proportion of obese European adults (31%). Māori had the second highest obesity prevalence (57%) in the same year and 43% of people who identified themselves as having multiple ethnicities with Māori ancestry were obese. If overweight status was included, more than 90% of Pacific and more than 80% of Māori would be considered overweight/obese using the universal BMI cut-offs over the survey periods (Figure 4.5). There was an overall increasing trend of adult obesity prevalence over time.

Similar to adults, Pacific and Māori children had a higher obesity prevalence compared to the other ethnic groups. Almost half of Pacific and one in every four Māori were obese in 2014/15. On the other hand, only 7% of Asians and one in eleven Europeans were obese in the same year. The obesity prevalence among Pacific children had grown from 34% in 2006/07 to 47% in 2014/15, whereas the prevalence had decreased among Asians and remained stable in other ethnic groups in the same period. Even though the obesity prevalence among Asians decreased from 9% in 2006/07 to 7% in 2014/15, the proportion of overweight Asian children increased; this equates to an overall increase in overweight/obesity prevalence among Asians in the same period. Furthermore, if overweight status was included, more than six out of ten Pacific and around half of Māori children would be considered overweight/obese throughout the survey periods (Figure 4.5).

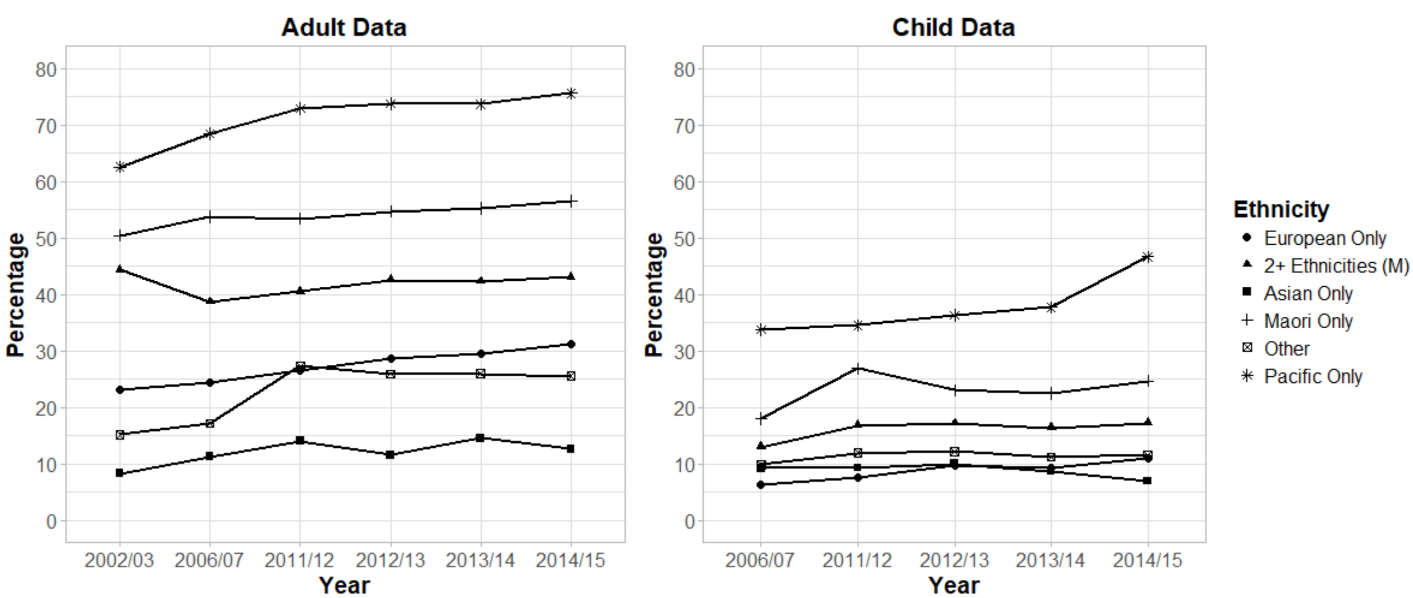

Figure 4.4: Age-Standardised Obesity Prevalence by Ethnicity from 2002/03 to 2014/15

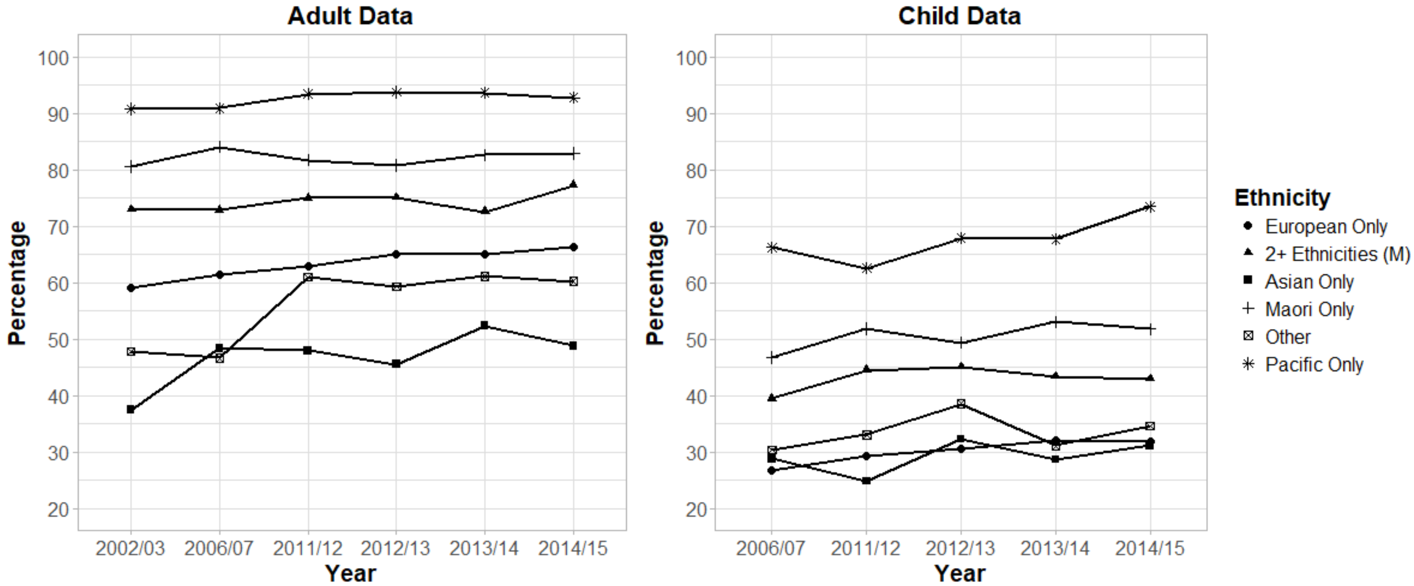

Figure 4.5: Age-Standardised Overweight/Obesity Prevalence by Ethnicity from 2002/03 to 2014/15

4.4 Regressions Result

I conducted a forward stepwise proportional odds regression with complex survey design in the Adult data to determine the final model. Before controlling for other covariates, people who lived in quintile five had 90% more risk of being in the higher BMI tertiles than those who lived in quintile one (OR = 1.89, 95%CI = 1.75–2.05). Linear contrast was then added to the model and was significant (OR = 1.21, 95%CI = 1.15–1.28); indicating that there was a secular change in BMI tertiles over time. Meeting vegetable guideline (OR = 1.03, 95%CI = 0.99–1.07) and fruit guideline (OR = 0.97, 95%CI = 0.93–1.005) did not exert a significant effect on BMI tertiles after adjusting for deprivation quintiles and linear contrast. On the same step, drinking problem was not a good predictor of BMI tertiles (OR = 0.99, 95%CI = 0.94–1.04). Sedentary lifestyle also lost its significance after adjusting for deprivation, linear contrast, ethnicity, age and difficulty climbing several flights of stairs (OR = 1.06, 95%CI = 0.99–1.12). There was no difference in BMI tertiles between people who lived in urban and rural areas after controlling for smoking status, sex, difficulty climbing stairs, age, ethnicity, and linear contrast (OR = 1.07, 95%CI = 0.98–1.17).

The summary of the final regression results are presented in Table 4.4. Having a higher educational qualification was protective against being in the higher BMI tertiles. Similarly, current smokers were less likely to be in the higher BMI tertiles, which indicates the 'nicotine effect'. All other variables in the model increased the likelihood of being in the higher BMI tertiles. The odds of being in the higher BMI tertiles increased as the level of area deprivation increased. Higher household income was associated with a higher risk of obesity, this finding is unusual considering a positive relationship between area level deprivation and obesity. Ethnicity remained a strong predictor of obesity even after adjusting for other variables.

I conducted another forward stepwise proportional odds regression with complex survey design to build a model for the Child data. The effect of meeting fruit guidelines (OR = 0.93, 95%CI = 0.84-1.03) and vegetable guidelines (OR = 0.99, 95%CI = 0.91–1.09) were not significant after adjusting for deprivation level and linear contrast. There was no evidence that urban/rural area had an influence on BMI categories after controlling for deprivation level and linear contrast (OR = 0.96, 95%CI = 0.8–1.14). Unlike the finding in adults, the BMI outcome of children who lived in the poorest household income ($\leq$\$15,000) was not significantly different than those in the richest household income (>\$70,000) after adjusting for ethnicity and deprivation level (OR = 1.23, 95%CI = 0.96-1.59). Fast food consumption was not a significant predictor of BMI outcome (4+/week versus <1/week; OR = 1.08, 95%CI = 0.83–1.39) after adjusting for soft drink consumption, education, age, ethnicity, and deprivation; this may suggest that there was a multicollinearity in the relationship between soft drink and fast food consumption.

The final regression results are presented in Table 4.5. Higher parents' educational qualification protected the children against overweight/obesity. Ethnicity remained a strong predictors of BMI outcomes after adjusting for demographic variables and soft drink consumption. The effect of ethnicity, however, was not as strong as in adults. Higher deprivation level equated to higher likelihood of being in the higher BMI categories in children, and the effect was stronger compared to the finding in adults.

Table 4.4: The Regression Summary of The Adult Data

	Odds Ratio	**2.50%**	**97.50%**
Deprivation Quintile (ref. group: Quintile 1)			
Quintile 2	1.02	0.94	1.11
Quintile 3	1.10	1.01	1.20
Quintile 4	1.29	1.19	1.40
Quintile 5	1.46	1.33	1.60
Linear Contrast	1.19	1.11	1.28
Ethnicity (ref. group: European only)			
2+ ethnicities (M)	1.72	1.57	1.88
Asian only	0.53	0.48	0.60
Māori only	3.01	2.75	3.29
Pacific only	8.41	7.22	9.79
Other	0.91	0.84	0.98
Age	1.01	1.01	1.01
Difficulty Climbing Several Flights of Stairs (ref. group: No difficulty)			
A little difficult	1.81	1.68	1.96
A lot difficult	1.80	1.64	1.98
Sex			
Female	0.76	0.73	0.80
Smoking Status (ref. group: Non smoker)			
Ex smoker	1.36	1.29	1.44
Current smoker	0.93	0.86	1.00
Household Income (ref. group: ≤$70,000)			
$15,001–$40,000	1.19	1.07	1.32
$40,001–$70,000	1.41	1.26	1.57
>$70,000	1.59	1.42	1.77
Educational Qualification (ref. group: No qualifacation)			
Secondary	0.89	0.83	0.96
Tertiary	0.80	0.74	0.86
Physical Activity			
Not physically active	1.15	1.09	1.20
Migration Status			
Native	1.22	1.10	1.34

Table 4.5: The Regression Summary of The Child Data

	Odds Ratio	**2.50%**	**97.50%**
Deprivation Quintile (ref. group: Quintile 1)			
Quintile 2	1.13	0.94	1.35
Quintile 3	1.41	1.20	1.67
Quintile 4	1.50	1.28	1.76
Quintile 5	1.76	1.50	2.06
Linear Contrast	1.16	1.05	1.28
Ethnicity (ref. group: European only)			
2+ ethnicities (M)	1.49	1.33	1.67
Asian only	1.04	0.84	1.28
Māori only	2.03	1.77	2.32
Pacific only	3.93	3.31	4.68
Other	1.21	1.05	1.38
Age	1.04	1.03	1.06
Parent's Qualification (ref. group: No qualification)			
Secondary	0.73	0.64	0.83
Tertiary	0.62	0.55	0.71
Soft Drink Consumption (ref. group: <1/week)			
1/week	1.15	1.03	1.29
2-3/week	1.24	1.10	1.39
4+/week	1.213	1.06	1.39
Sex			
Female	1.11	1.02	1.21

4.5 Proportional Odds Assumption

The proportional odds assumption in the Adult data did not hold for sex, where male adults were more likely to belong in the mid or top tertile while female adults were more likely to belong in the top tertile. This relationship can also be seen in the sex by BMI tertiles chart (first row and second column of Figure 4.2), this showed that there were more female in the top tertile and there were more male in the mid tertile. Therefore, interpretation of the sex variable of the proportional odds logistic regression should be done carefully. Furthermore, adults who had difficulties climbing several flights of stairs were more likely to be in the top tertile compared with being in the mid or top tertile. In Pacific adults, the likelihood to be in the mid or top tertile was higher compared with being in the top tertile. In children, the proportional odds assumption did not hold for quintile three to five, Māori and Pacific ethnic group; on all of those variables, the likelihoods of being in the obese category were higher compared with the likelihoods of being in the overweight/obese category. The details of this analyses are presented in Listing A.2 and the binomial regression tables for assessing the proportional odds assumption are presented in Appendix C.

5 Discussion

5.1 Summary of the findings

This study found that the prevalence of obesity in adults and children had not changed significantly since 2011/12. Despite that, I found that there was evidence of linear temporal change over time from the regression analysis. The regression analysis showed that ethnicity was a strong predictor of BMI outcomes. This further asserts the finding from the descriptive statistics showing that Pacific and Māori had the highest obesity prevalence among other ethnic groups. The difference in BMI outcomes between ethnicities was even more obvious in adults, suggesting that the BMI across ethnic groups will keep diverging as people age. The main objective of this study is to assess whether socio-economic deprivation, as measured by area level deprivation, influences the BMI outcome. This study found that the odds of being in the higher BMI groups increased as the level of area deprivation where people reside increased, and the association was stronger in children.

In the next section, I will address the association between the area level deprivation and obesity followed by a discussion regarding an accidental finding from this study about household income and obesity. I will then discuss BMI differences across ethnic groups and the problem of using the universal BMI cut-offs. I will also talk about the findings related to diets and physical activity. Lastly, I will discuss the smoking habit, healthy immigrant effect and living in urban/rural areas in relation to obesity.

5.2 Area-Level Deprivation

This study found that the deprivation quintile influenced one's body mass index after adjusting for household income, age, education level, and ethnicity. This finding is consistent with other studies in the US [Nau et al., 2015], Canada [Matheson et al., 2008], UK [Cummins and Macintyre, 2006] and Sweden [Stafford et al., 2010]. Adults and children who lived in the most deprived communities had a higher trajectory of BMI growth over time [Stafford et al., 2010, Nau et al., 2015]. Moreover, studies have found that people who lived in more deprived communities had poorer access to fruit and vegetables [Eagle et al., 2012, Cummins and Macintyre, 2006, Nau et al., 2015], which would lead to a poorer diet quality [Craig et al., 2016]. Furthermore, children in deprived neighbourhoods had lower opportunities for outdoor play due to inadequate infrastructure as well as safety concerns [Eagle et al., 2012, Nau et al., 2015]. On top of that, structural and social barriers to health care services in more deprived communities put people at a higher risk of being overweight/obese [Li et al., 2014]. This study highlights that even in a country with a strong welfare system, people who lived in more deprived areas had higher odds of developing overweight/obesity. It does not seem that poor people become more obese because they are being "lazy", instead a lot of social and infrastructure barriers put them in a more disadvantaged position with less resources to combat obesity.

5.3 Household Income Level

The association between socio-economic status and obesity seems to depend on the income level of the country. Socio-economic status was inversely related to obesity rates among high-income countries, but the association was reversed in low-income countries [Wu et al., 2015, Fruhstorfer et al., 2016]. In high-income countries, food insecurity is not prevalent anymore even among low-income families due to a strong welfare system that helps them with foods and daily needs. Nonetheless, they are more susceptible to obesity compared with their well-off counterparts due to a lack of resources to choose a healthier diet and lifestyle, which tend to be more expensive. Conversely, in low-income countries, the association was reversed. High-income families were more likely to be overweight/obese than the low-income families. This association occurred because food availability is still a challenge in low-income families and overweight may be perceived as a sign of wealth [Dinsa et al., 2012, Wu et al., 2015, Fruhstorfer et al., 2016].

In New Zealand, contrary to what would have been expected from a high-income country, this study found that household income was positively associated with adult BMI tertiles. It was found that low-income families spent less than the estimated cost required for a basic diet. Furthermore, in opposition to the popular belief with respect to eating behaviours in low-income groups, these families only spent 5% of the total food expenditure on ready-to-eat food [Smith et al., 2013, Smith et al., 2010]. This might suggest that adults from low-income families still experienced food insecurity which might have protected them from obesity. However, I did not find a statistically significant effect of household income on overweight/obesity in children. The Working for Family program, which pays extra money to almost all families with children who earned less than $57,000 per year, may explain the non-significant finding in children. This program ensured that low-income families with children would have enough money to afford food and their basic daily needs to sustain life. Further research on food securities among low-income families is warranted to shed more light on this issue, especially among those who do not receive any government benefit.

5.4 Obesity Trend

According to the OECD report in 2014, New Zealand is ranked third for its obesity prevalence; behind The United States and Mexico [OECD, 2017]. However, the figures reported in the OECD report did not account for the difference in ethnic group composition and it would have not provided a correct picture of the obesity epidemic. For example, 61% of the US population is White, 12% is Black, and 18% is Hispanic in 2016 [Kaiser Family Foundation, 2016]; and their obesity prevalence was 47.8% in Non-Hispanic Black and 42.5% in Hispanic, while it was only 32.6% among Non-Hispanic White [Ogden et al., 2013]. The figure presented in the OECD report for the US must have been influenced by the high rate of obesity among Non-Hispanic Black and Hispanic group. Similarly, in New Zealand, 74% of the population is European, 15% is Māori, and 7% is Pacific [Statistics New Zealand, 2017] in 2013; and this study found that 55% Māori and 74% Pacific adults were obese in 2012/13.

By contrast, the United Kingdom, which had a lower national obesity prevalence according to OECD report, consists of 86% White and 2.5% Indian [Office for National Statistics, 2011]. This would translate to a relatively lower national obesity prevalence compared with countries with a high proportion of ethnic groups who have predispositions to accumulate more body weight. Using universal obesity cut-offs across different ethnicities will result in lower obesity prevalence among Asian and White European people while it will be higher for Black African [El-Sayed et al., 2011], Non-Hispanic Black, Hispanic [Ogden et al., 2013], Māori, and Pacific people. Comparing the obesity prevalence across countries while disregarding each country's unique ethnic composition will not give an accurate picture of the problem. It is better to report obesity prevalence on each ethnic group when trying to make a comparison of the obesity epidemic across different countries.

5.5 Ethnic Differences and The Universal BMI Cut-offs

The effect of ethnicities on BMI tertiles or categories persisted even after controlling for other demographic variables. In this study, Asians (Indian and/or Chinese) were less likely to be in the higher BMI tertiles compared with Europeans. This finding agrees with other studies [Stommel and Schoenborn, 2010, El-Sayed et al., 2011, Barba et al., 2004], and the trend is also observed in other Asian ethnic groups such as Indians, Chinese, Malaysians [Deurenberg-Yap et al., 2002, Deurenberg-Yap et al., 2000], Indonesians, Japanese [Barba et al., 2004], and Polynesians [Deurenberg, 2001]. Despite the tendency of having a lower average BMI, Asians have a higher risk of developing type II diabetes and cardiovascular disease compared with Europeans at the same BMI level. On average, South Asians are 3.4 times more likely to develop diabetes compared with Europeans after adjusting for age, sex, baseline BMI and other sociodemographic factors [Chiu et al., 2011]. This tells us that a higher BMI level do not always mean a higher risk of developing diabetes when looking at a population with diverse ethnic composition.

Compared with Caucasians at the same BMI level; Indians, Chinese, and Malays have a higher body fat percentage [Deurenberg-Yap et al., 2002, Deurenberg-Yap et al., 2000]. Despite that, Indians were still more likely to have insulin resistance compared with Caucasians even after controlling for diet, anthropometric measurement, and body fat percentage [Abate and Chandalia, 2001]. Other ethnic groups such as American Indians, African American, Pacific Islanders [Dabelea et al., 2007], Chinese, Japanese, Malaysians also had a higher incidence of insulin resistance and type II diabetes compared with Europeans [Gao et al., 2012]. It may be possible that Asians may have a genetic predisposition to develop insulin resistance [Abate and Chandalia, 2001]. This association was further supported by a cohort study in Canada. It found that some ethnic groups develop type II diabetes at a much younger age than those of European descents after controlling for demographic variables and BMI. On average, diabetes presented itself nine years earlier in South Asian descent, three years earlier in Chinese descent, and one year earlier among Black descent compared with European descents [Chiu et al., 2011]. These suggest that BMI is a poor predictor of diabetes risk across diverse ethnic groups and there may be a genetic influence at play.

Despite the fact that Pacific and Māori people had a lower body fat percentage compared with other ethnic groups at the same BMI level [Rush et al., 2009], diabetes prevalence among Pacific and Māori remained high relative to the rest of New Zealand population [Hawley and McGarvey, 2015]. This study also found that Pacific adults were eight times and Pacific children were almost four times more likely to be in the higher BMI groups compared with Europeans. The same association was also found in Māori, where Māori adults and children were three and two times more likely to be in the higher BMI groups compared with Europeans. Furthermore, by using the universal BMI cut-offs, more than 70% of Pacific adults and more than half of Māori adults would be identified as obese. If overweight status was included, more than 90% and 80% of Pacific and Māori adults would be identified as either overweight or obese. This further questions the usefulness of the "average" BMI category and more research is needed in order to determine whether existing overweight/obesity cut-offs are accurate in predicting diabetes risk in these groups.

5.6 Diet and Physical Activity

5.6.1 Fruit and Vegetable Guidelines

I found no evidence that adherence to fruit and vegetable guidelines influence the BMI outcome after adjusting for deprivation level. A systematic review have found that a reduction in overweight/obesity likelihood is only achieved when the increase in fruit and vegetable consumption is accompanied by a reduction in the total caloric intake [Ledoux et al., 2011]. Another longitudinal study in children also found that fruit and vegetable consumption was not associated with less energy-dense food consumption (e.g. high-energy drinks, sweets, and snacks); which meant the children would have the same, if not more, total caloric intake than those who consumed less fruit and vegetable [Bayer et al., 2014]. The result from this study may had happened because people who did consume more fruit and vegetables had the same total caloric intake as those who did not.

5.6.2 Fast Food and Soft Drink Consumption in Children

I found that the effect of soft drink consumption on childhood overweight/obesity remained significant even after adjusting for parent's educational qualification, age, ethnicity, and deprivation level. The association between soft drink or SSB consumption on weight gain is consistent across studies [Vartanian et al., 2007, De Ruyter et al., 2012] even after adjusting for various diets and lifestyle factors [Malik et al., 2013a]. Interestingly, the association was attenuated when total energy intake was taken into account. Even so, soft drink consumption is consistently associated with increased total energy intake. This suggests that people who consumed SSB might not have been able to compensate for the extra energy [Vartanian et al., 2007]. Following administration of 500ml of either water or SSB, people would still consume the same amount of energy from food. This would result in a higher total energy intake for those who drank SSB compared with those who only drank water [Maersk et al., 2012]. Even though the effect of SSB consumption is attenuated by the total caloric intake, it is clear that those who drank SSB did not reduce their total caloric intake, this consequently put them in a higher risk of being obese.

On the other hand, I found that the frequency of fast food consumption was not associated with overweight/obesity status in children. This result agrees with a study in China regarding child obesity and fast food consumption. Even though there was a significant increase in fast food consumption among adolescents from 2004 to 2009; it was not accompanied by any change in BMI z-score after adjusting for age, ethnicity, household income, geographical region, and physical activity level [Xue et al., 2016]. Overall, the associations between fast food consumption and obesity were not found in children, but it is consistently significant in adults. A Possible explanation for this is that children require more energy while growing up, therefore extra caloric intake from fast food is not readily converted into body fat [Rosenheck, 2008]. Despite that, because fast food consumption was associated with soft drink consumption, it is possible that children who consumed fast food would increase their soft drink intake and in turn lead to a higher likelihood of being obese. The non-significant finding from this study might be a consequence of multi-colinearity between soft drink and fast food consumption.

5.6.3 Physical Activity

This study found that sedentary lifestyle (spending less than 30 minutes of exercise in the past week) did not have a significant influence on the BMI tertiles after adjusting for difficulty climbing several flights of stairs, which is a proxy for physical capability. It is possible that people who were sedentary had medical conditions that might lower their capability to exercise and make them more prone to developing obesity. For example, obese people have a higher risk of developing knee problems [Banjare and Bhalerao, 2016], which in turn lead to an increase in sedentary activity [Joffe, 2016] and make it difficult for them to reverse their obesity. On the other hand, people who were physically active (spent at least 30 minutes of exercise on five or more days in the past week) had a lower likelihood of being in the higher BMI tertiles. Nevertheless, because reverse causation cannot be ruled out in this type of study and the evidence have shown that physical activity intervention was not effective in maintaining long-term weight loss [Kovacs et al., 2014, Shaw et al., 2006], it is unlikely that being physically active would reduce the risk of obesity. Instead, this finding may indicate that people with less functional limitations, which had normal weight, were more likely to be physically active.

5.7 Smoking Status

This study found that current smokers had a lower likelihood of being in the higher BMI tertiles compared with the non-smokers. In contrast, ex-smokers had a higher risk of being in the higher BMI tertiles. These findings are consistent with other studies where people tend to gain more weight after smoking cessation [MacKay et al., 2013, Rom et al., 2015, Siahpush et al., 2014]. The reduction in body weight among smokers is expected because nicotine activates the sympathetic nervous system and increases energy expenditure [Rigotti and Clair, 2013]. The weight gain after smoking cessation is observed as early as one year after quitting, but only in older adults. However, the weight gain was not observed in people aged 16–24 years who quit smoking [MacKay et al., 2013]. Although the weight gained is mostly attributable to an increase in body fat; a study showed that ex-smokers also had higher muscle mass, muscle strength, and bone density compared with current smokers [Rom et al., 2015]. Additionally, overweight or obese ex-smokers have a lower overall mortality risk compared with current smokers with normal weight [Siahpush et al., 2014]. Therefore, there is no benefit in preventing people who want to quit smoking in order to stay lean because continuing smoking even with normal weight would put people in a higher risk of developing adverse health events. It is noteworthy that an intervention focusing on one aspect of health (i.e. body weight) may not be appropriate for some people and careful considerations are needed when engaging overweight/obese individuals with unique problems.

5.8 Healthy Immigrant

This study found that adults who had always stayed in New Zealand and those who had migrated to New Zealand for more than 10–11 years ago were more likely to be in the higher BMI tertiles compared with those who migrated to New Zealand less than 10–11 years ago. The protective effect of being a migrant remained significant regardless of their ethnicity, age, education level, and sex. This finding is similar to studies from other countries such as the US, Canada, Australia, and the United Kingdom [Vang et al., 2015, Kennedy et al., 2015, Goel et al., 2004]. Better BMI outcome among the immigrants might be explained by the immigration selection favoring healthy individuals [Kennedy et al., 2015]. However, it is very likely that the protective effect will diminish over time, and the health risk of immigrants will match the risk of the host country after 10–15 years [Vang et al., 2015, Ledoux et al., 2011, Goel et al., 2004].

Interestingly, another study did not found the healthy immigrant effect among Pacific people in New Zealand. It was found that not only Pacific people had received preference over other ethnic groups, but they also came from a more health disadvantaged communities [Hajat et al., 2010]. This study did not conduct a sub-group analysis, thus could not draw a specific conclusion regarding healthy immigrant effect among Pacific people. In light of that, interpreting the healthy immigrant effect should be done carefully as it might not apply to some groups of people due to other circumstances that were not captured in this study.

5.9 Urbanisation

Urbanisation has been thought to cause the obesity epidemic through various mechanisms [Kearney, 2010, Malik et al., 2013b, Sturm and An, 2014, Durand et al., 2011], but the evidence is conflicting. A study in India found that rural people who migrated to urban areas had a higher BMI and diabetes prevalence compared with rural dwellers. Those who migrated to urban areas had an increased fat intake and reduced physical activity, which may explain higher levels of obesity and diabetes among this group [Ebrahim et al., 2010]. However, a study in China found that people who lived in rural areas were more likely to be overweight/obese compared with those in urban areas. It appears that China has managed to improve food security even in rural settings, which lead to higher caloric consumption. Moreover, due to agricultural mechanisation, a lot more Chinese people in rural areas did not exercise as often as people in metropolitan areas [Tian et al., 2014]. This study found that there was no BMI difference between urban and rural dwellers. This finding suggests that the urban/rural indicator may not be a reliable predictor of overweight/obesity status. There may be more important demographic information such as deprivation level and ethnicity, which have a stronger influence on BMI.

5.10 Study Strengths and Limitations

To my knowledge, this is the first study in New Zealand that uses panels of cross-sectional surveys of individual data to assess the association between deprivation level and BMI. The analysis was done using the NZHS data, which contains information on various demographic variables and health-related behaviours. In that regard, the association between deprivation level and BMI outcome in this study is reflective of the true effect estimate as it is adjusted for possible confounding factors such as sex, age, household income, smoking status, physical activity level, and diets. This study also used a complex survey design, which took into account the sampling design used in the NZHSs. This ensured the result of this analysis is generalisable to the New Zealand population.

Despite the use of multiple cross-sectional surveys over time, findings in this study is not free from the possibilities of reverse causation. This study also did not have access to individual identifiers, and thus are unable to exclude the possibility that the same person could be measured more than once throughout the survey periods. Moreover, the information on deprivation quintile in NZHS 2011/12 and 2012/13 came from the census data in 2006; this study assumed that the area level deprivation had not changed significantly as to have an impact on the BMI outcome at those survey periods. The NZHS also did not have information on migrations within New Zealand, it is possible that healthier people moved into communities with a lower deprivation level which might have confounded the findings. This study also did not conduct subgroup analyses across ethnic groups. Considering that more than 80% of Pacific and Māori belong to the higher BMI categories, it is possible that the relationship between area level deprivation and the BMI outcomes might change for a certain group.

6 Conclusions

6.1 Implications of Socio-Economic Status and BMI outcomes

Higher levels of area deprivation put people at a higher risk of developing obesity. This finding persists even after adjusting for other demographic factors and health-related behaviours. Lack of socio-economic and health care resources in this area makes it harder for people to prevent and/or reverse obesity. Imposing behavioural changes on these people should be done with consideration, as they might not have the resources (i.e. time, money, facilities, etc.) to modify their habits. In order to reduce the disparity in the obesity prevalence across different area deprivation levels, the government should devise a population-level intervention that improves the socio-economic status of the area and also create an environment that is conducive to maintaining a healthy lifestyle. This study also found that, unlike other high-income countries, high-income families in New Zealand had a higher risk of developing obesity. This may indicate that low-income families still experienced food security issues in their everyday lives.

6.2 Implications of The Universal BMI Cut-Offs

The intention of categorising people into different BMI groups is to identify the health risks associated with a certain BMI group and to devise appropriate interventions that may mitigate the associated health risks. However, as I have discussed, universal BMI cut-offs are not accurate in assessing the obesity-associated health risks in a population with diverse ethnic groups. Different ethnic groups have different average body weight, and their risk of developing diabetes is different at the same BMI value. We found that the obesity prevalence in Asians was just under 15% and despite having a much lower average BMI, their risk of developing diabetes is higher compared with the other ethnic groups at the same BMI level. Using BMI alone as an indicator of population health would mean that Asians will receive little attention. On the other hand, a higher average BMI among Pacific and Māori would mean that they will receive rigorous interventions which lead to another problem. If Pacific and Māori have a higher average BMI, at what BMI point can the intervention be said to have achieved its goal? Further imposing more strenuous diet and/or exercise interventions would only create frustration without achieving the actual goals (i.e. reducing the risk of obesity-associated comorbidities). Forcing the use of the "average" BMI category as a blanket public health goal may harm certain ethnic groups.

6.3 Implications of Diet and Physical Activity Interventions

This study found that lower soft drink consumption is associated with lower likelihood to be in the higher BMI categories. However, no association is found for fast food consumption as well as adherence to fruit and vegetable guidelines. This, however, does not mean consuming more fast food will not have any impact on BMI. Due to correlation between fast food and soft drink consumption, it is likely that people who consume more fast food will increase their soft drink intake as well; this in turn will increase the risk of developing obesity. Replacing fast food and soft drink with more fruit and vegetables, despite having no effect on BMI, would fulfill the daily nutritional and micro-nutrients requirement. It is important to note that improving population dietary habits with more fruit and vegetables will benefit health regardless of weight change, and no weight change should not be used as an argument to deter from eating healthier foods.

The result from this study also demonstrates that there was no difference in obesity risk between people who had a sedentary and non-sedentary lifestyle. The benefit of physical activity on reducing the likelihood of being in the higher BMI groups was only realised when people spent more than 30 minutes of moderate physical activity or brisk walking on at least five days in a week. This criteria appears to be unrealistic because only a small fraction of the population can achieve this. Moreover, it is possible that those who did not engage in physical activity might have had other medical problems that limit their functional capacity. Standard physical activity interventions may not be suitable for this group of people, and imposing a desired BMI on them is highly inappropriate. As have been shown from multiple studies, physical activity interventions did not manage to maintain long-term weight loss thus the drive to increase physical activity level should not be based on weight reduction. Increasing physical activity whenever and wherever appropriate will benefit other aspects of health (e.g. improved blood pressure, lower lipid profiles and fasting serum glucose) regardless of weight change.

6.4 Recommendations

People living in more deprived areas are faced with socio-economic barriers and structural challenges in achieving a healthier lifestyle. Therefore, devising public health interventions and/or health policies to reduce obesity should not only empower them with knowledge about nutrition and exercise but also improve infrastructure support (e.g., more accessible and healthier food options, better access to health care, etc.).

Furthermore, reliance on BMI or other body weight measurements as the only population or individual health goal should stop because different ethnic groups have a different risk of developing adverse health conditions at the same BMI level. Instead, dietary habits and physical activity levels before and after an intervention can be used as an indicator of successful progress towards a healthier life regardless of weight change.

Longitudinal research assessing the impact of area level deprivation on obesity-associated comorbidities in New Zealand is needed. This is important because obesity on its own is not the main problem, whereas the diseases associated with obesity are the conditions that impact on quality of life and increase health care costs. Also, more research is needed to assess the risk of developing obesity-associated chronic conditions among Pacific and Māori people in New Zealand. This will also help determine whether specific overweight/obesity cut-offs or a better indicator of population health other than weight measurements are needed for these groups.

Bibliography

[Abate and Chandalia, 2001] Abate, N. and Chandalia, M. (2001). Ethnicity and type 2 diabetes: Focus on Asian Indians. *Journal of Diabetes and its Complications*, 15(6):320–327.

[Abdullah et al., 2011] Abdullah, A., Stoelwinder, J., Shortreed, S., Wolfe, R., Stevenson, C., Walls, H., de Courten, M., and Peeters, A. (2011). The duration of obesity and the risk of type 2 diabetes. *Public Health Nutrition*, 14(1):119–126.

[Agresti, 2013] Agresti, A. (2013). *Categorical data analysis*, volume 792. Wiley, Hoboken, NJ, 3rd edition.

[Alexandratos and Bruinsma, 2012] Alexandratos, N. and Bruinsma, J. (2012). World agriculture towards 2030/2050: The 2012 revision. Report, ESA Working paper Rome, FAO.

[Angkurawaranon et al., 2014] Angkurawaranon, C., Jiraporncharoen, W., Chenthanakij, B., Doyle, P., and Nitsch, D. (2014). Urban environments and obesity in Southeast Asia: a systematic review, meta-analysis and meta-regression. *PLoS ONE*, 9(11):1–19.

[Antonini et al., 2007] Antonini, L., Grosso, A., Francalanci, C., and Mattei, R. (2007). Cardiovascular pathophysiology of obesity. *High Blood Pressure & Cardiovascular Prevention*, 14(4):207–212.

[Antonopoulos et al., 2016] Antonopoulos, A. S., Oikonomou, E. K., Antoniades, C., and Tousoulis, D. (2016). From the BMI paradox to the obesity paradox: the obesity-mortality association in coronary heart disease: obesity paradox in CHD. *Obesity Reviews*, 17(10):989–1000.

[Atkinson et al., 2014] Atkinson, J., Salmond, C., and Crampton, P. (2014). *NZDep2013 Index of Deprivation*. University of Otago, Dunedin.

[Banjare and Bhalerao, 2016] Banjare, J. B. and Bhalerao, S. (2016). Obesity associated noncommunicable disease burden. *International Journal of Health & Allied Sciences*, 5(2):81–87.

[Barba et al., 2004] Barba, C., Cavalli-Sforza, T., Cutter, J., Darnton-Hill, I., Deurenberg, P., Deurenberg-Yap, M., Gill, T., James, P., Ko, G., Nishida, C., and Consultation, W. H. O. E. (2004). Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *The Lancet*, 363(9403):157–163.

[Barness et al., 2007] Barness, L. A., Opitz, J. M., and Gilbert-Barness, E. (2007). Obesity: genetic, molecular, and environmental aspects. *American Journal of Medical Genetics Part A*, 143A(24):3016–3034.

[Bayer et al., 2014] Bayer, O., Nehring, I., Bolte, G., von Kries, R., and Cohort, G. (2014). Fruit and vegetable consumption and BMI change in primary school-age children: A cohort study. *European Journal of Clinical Nutrition*, 68(2):265–270.

[Bogardus, 2009] Bogardus, C. (2009). Missing heritability and GWAS utility. *Obesity*, 17(2):209–210.

[Buchmueller and Johar, 2015] Buchmueller, T. C. and Johar, M. (2015). Obesity and health expenditures: Evidence from Australia. *Economics & Human Biology*, 17:42–58.

[Buchwald et al., 2004] Buchwald, H., Avidor, Y., Braunwald, E., Jensen, M. D., Pories, W., Fahrbach, K., and Schoelles, K. (2004). Bariatric surgery: A systematic review and meta-analysis. *JAMA*, 292(14):1724–1737.

[Cali and Caprio, 2008] Cali, A. M. G. and Caprio, S. (2008). Obesity in children and adolescents. *The Journal of Clinical Endocrinology & Metabolism*, 93(11_supplement_1):s31–s36.

[Centers for Disease Control and Prevention, 2016] Centers for Disease Control and Prevention (16 June 2016). Defining adult overweight and obesity. `http://www.cdc.gov/obesity/adult/defining.html`. Accessed: 22 April 2017.

[Centers for Disease Control and Prevention, 2013] Centers for Disease Control and Prevention (May 2013). Use and interpretation of the WHO and CDC growth chart for children from birth to 20 years in the united states. `https://www.cdc.gov/nccdphp/dnpa/growthcharts/resources/growthchart.pdf`. Accessed: 22 April 2017.

[Chaparro et al., 2014] Chaparro, M. P., Whaley, S. E., Crespi, C. M., Koleilat, M., Nobari, T. Z., Seto, E., and Wang, M. C. (2014). Influences of the neighbourhood food environment on adiposity of low-income preschool-aged children in los angeles county: a longitudinal study. *Journal of Epidemiology and Community Health*, 68(11):1027–1033.

[Chiu et al., 2011] Chiu, M., Austin, P. C., Manuel, D. G., Shah, B. R., and Tu, J. V. (2011). Deriving ethnic-specific BMI cutoff points for assessing diabetes risk. *Diabetes Care*, 34(8):1741–1748.

[Cohen et al., 2013] Cohen, A. K., Rai, M., Rehkopf, D. H., and Abrams, B. (2013). Educational attainment and obesity: A systematic review. *Obesity Reviews*, 14(12):989–1005.

[Cohen, 1992] Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1):155–159.

[Cole et al., 2000] Cole, T. J., Bellizzi, M. C., Flegal, K. M., and Dietz, W. H. (2000). Establishing a standard definition for child overweight and obesity worldwide: International survey. *British Medical Journal*, 320(7244):1240–1243.

[Cole and Lobstein, 2012] Cole, T. J. and Lobstein, T. (2012). Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity. *Pediatric Obesity*, 7(4):284–294.

[coreysparks, 2015] coreysparks (2015). DEM 7283 — Example 4 — ordinal & multinomial logit models. `https://rstudio-pubs-static.s3.amazonaws.com/58926_5f62fd1d242745c4a2a28b18376118df.html`. Accessed: 10 April 2017.

[Cotti and Tefft, 2013] Cotti, C. and Tefft, N. (2013). Fast food prices, obesity, and the minimum wage. *Economics and Human Biology*, 11(2):134–147.

[Craig et al., 2016] Craig, L. C. A., McNeill, G., Masson, L. F., and Macdiarmid, J. I. (2016). Diet quality index in children in scotland: associations with age, sex, socio-economic deprivation and obesity. *Proceedings of the Nutrition Society*, 75(OCE3).

[Crombie et al., 2008] Crombie, I. K., Kiezebrink, K., Irvine, L., Wrieden, W. L., Swanson, V., Power, K., and Slane, P. W. (2009;2008;). What maternal factors influence the diet of 2-year-old children living in deprived areas? a cross-sectional survey. *Public Health Nutrition*, 12(8):1254–1260.

[Cummins and Macintyre, 2006] Cummins, S. and Macintyre, S. (2006). Food environments and obesity — neighbourhood or nation? *International Journal of Epidemiology*, 35(1):100–104.

[Dabelea et al., 2007] Dabelea, D., Bell, R. A., D'Agostino Jr, R. B., Imperatore, G., Johansen, J. M., Linder, B., Liu, L. L., Loots, B., Marcovina, S., Mayer-Davis, E. J., et al. (2007). Incidence of diabetes in youth in the united states. *JAMA*, 297(24):2716–2724.

[De Ruyter et al., 2012] De Ruyter, J. C., Olthof, M. R., Seidell, J. C., and Katan, M. B. (2012). A trial of sugar-free or sugar-sweetened beverages and body weight in children. *New England Journal of Medicine*, 367(15):1397–1406.

[de Vos et al., 2016] de Vos, B. C., Runhaar, J., van Middelkoop, M., Krul, M., and Bierma-Zeinstra, S. M. (2016). Long-term effects of a randomized, controlled, tailor-made weight-loss intervention in primary care on the health and lifestyle of overweight and obese women. *Am J Clin Nutr*, 104(1):33–40.

[Deurenberg, 2001] Deurenberg, P. (2001). Universal cut-off BMI points for obesity are not appropriate. *British Journal of Nutrition*, 85(2):135–136.

[Deurenberg-Yap et al., 2002] Deurenberg-Yap, M., Chew, S., and Deurenberg, P. (2002). Elevated body fat percentage and cardiovascular risks at low body mass index levels among singaporean chinese, malays and indians. *Obesity Reviews*, 3(3):209–215.

[Deurenberg-Yap et al., 2000] Deurenberg-Yap, M., Schmidt, G., Staveren, van, W., and Deurenberg, P. (2000). The paradox of low body mass index and high body fat percentage among Chinese, Malays and Indians in singapore. *International Journal of Obesity*, 24(8):1011–1017.

[Didham and Callister, 2012] Didham, R. and Callister, P. (2012). The effect of ethnic prioritisation on ethnic health analysis: a research note. *The New Zealand medical journal*, 125(1359):58–66.

[Dijkstra et al., 2015] Dijkstra, S. C., Neter, J. E., van Stralen, M. M., Knol, D. L., Brouwer, I. A., Huisman, M., and Visser, M. (2015). The role of perceived barriers in explaining socio-economic status differences in adherence to the fruit, vegetable and fish guidelines in older adults: a mediation study. *Public Health Nutrition*, 18(5):797–808.

[Dinsa et al., 2012] Dinsa, G. D., Goryakin, Y., Fumagalli, E., and Suhrcke, M. (2012). Obesity and socioeconomic status in developing countries: a systematic review. *Obesity Reviews*, 13(11):1067–1079.

[Dugas et al., 2011] Dugas, L. R., Harders, R., Merrill, S., Ebersole, K., Shoham, D. A., Rush, E. C., Assah, F. K., Forrester, T., Durazo-Arvizu, R. A., and Luke, A. (2011). Energy expenditure in adults living in developing compared with industrialized countries: a meta-analysis of doubly labeled water studies. *The American journal of clinical nutrition*, 93(2):427–441.

[Durand et al., 2011] Durand, C. P., Andalib, M., Dunton, G. F., Wolch, J., and Pentz, M. A. (2011). A systematic review of built environment factors related to physical activity and obesity risk: implications for smart growth urban planning. *Obesity Reviews*, 12(501):e173–e182.

[Eagle et al., 2012] Eagle, T. F., Sheetz, A., Gurm, R., Woodward, A. C., Kline-Rogers, E., Leibowitz, R., Durussel-Weston, J., Palma-Davis, L., Aaronson, S., Fitzgerald, C. M., Mitchell, L. R., Rogers, B., Bruenger, P., Skala, K. A., Goldberg, C., Jackson, E. A., Erickson, S. R., and Eagle, K. A. (2012). Understanding childhood obesity in america: Linkages between household income, community resources, and children's behaviors. *American Heart Journal*, 163(5):836–843.

[Ebrahim et al., 2010] Ebrahim, S., Kinra, S., Bowen, L., Andersen, E., Ben-Shlomo, Y., Lyngdoh, T., Ramakrishnan, L., Ahuja, R. C., Joshi, P., Das, S. M., Mohan, M., Smith, G. D., Prabhakaran, D., Srinath Reddy, K., Grp, I. M. S., group, I. M. S., and for the Indian Migration Study group (2010). The effect of rural-to-urban migration on obesity and diabetes in India: a cross-sectional study. *PLoS Medicine*, 7(4):e1000268.

[El-Sayed et al., 2011] El-Sayed, A. M., Scarborough, P., and Galea, S. (2011). Ethnic inequalities in obesity among children and adults in the UK: a systematic review of the literature. *Obesity Reviews*, 12(501):e516–e534.

[Evans et al., 2000] Evans, J. M. M., Newton, R. W., Ruta, D. A., MacDonald, T. M., and Morris, A. D. (2000). Socio-economic status, obesity and prevalence of type 1 and type 2 diabetes mellitus. *Diabetic Medicine*, 17(6):478–480.

[Exeter et al., 2015] Exeter, D., Sabel, C., Hanham, G., Lee, A., and Wells, S. (2015). Movers and stayers: the geography of residential mobility and CVD hospitalisations in Auckland, New Zealand. *Social Science & Medicine*, 133:331–339.

[Flegal et al., 2013] Flegal, K. M., Kit, B. K., Orpana, H., and Graubard, B. I. (2013). Association of all-cause mortality with overweight and obesity using standard body mass index categories: a systematic review and meta-analysis. *JAMA*, 309(1):71–82.

[Fruhstorfer et al., 2016] Fruhstorfer, B. H., Mousoulis, C., Uthman, O. A., and Robertson, W. (2016). Socio-economic status and overweight or obesity among school-age children in sub-saharan africa — a systematic review: obesity among children in sub-Saharan Africa. *Clinical Obesity*, 6(1):19–32.

[Gao et al., 2012] Gao, H., Salim, A., Lee, J., Tai, E. S., and Van Dam, R. M. (2012). Can body fat distribution, adiponectin levels and inflammation explain differences in insulin resistance between ethnic Chinese, Malays and Asian Indians. *International Journal of Obesity*, 36(8):1086–1093.

[Giskes et al., 2011] Giskes, K., van Lenthe, F., Avendano-Pabon, M., and Brug, J. (2011). A systematic review of environmental factors and obesogenic dietary intakes among adults: are we getting closer to understanding obesogenic environments? *Obesity reviews*, 12(501):e95–e106.

[Goel et al., 2004] Goel, M. S., McCarthy, E. P., Phillips, R. S., and Wee, C. C. (2004). Obesity among US immigrant subgroups by duration of residence. *Jama*, 292(23):2860–2867.

[Guh et al., 2009] Guh, D. P., Zhang, W., Bansback, N., Amarsi, Z., Birmingham, C. L., and Anis, A. H. (2009). The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis. *BMC Public Health*, 9(1):88–88.

[Hajat et al., 2010] Hajat, A., Blakely, T., Dayal, S., and Jatrana, S. (2010). Do New Zealand's immigrants have a mortality advantage? evidence from the New Zealand Census-Mortality Study. *Ethnicity & Health*, 15(5):531–547.

[Hawley and McGarvey, 2015] Hawley, N. L. and McGarvey, S. T. (2015). Obesity and diabetes in Pacific Islanders: the current burden and the need for urgent action. *Current Diabetes Reports*, 15(5):1–10.

[Herrera and Lindgren, 2010] Herrera, B. M. and Lindgren, C. M. (2010). The genetics of obesity. *Current Diabetes Reports*, 10(6):498–505.

[Hodgkin et al., 2010] Hodgkin, E., Hamlin, M. J., Ross, J. J., and Peters, F. (2010). Obesity, energy intake and physical activity in rural and urban New Zealand children. *Rural and remote health*, 10(2):1336.

[Hu et al., 2014] Hu, Y., Bhupathiraju, S. N., Koning, L., and Hu, F. B. (2014). Duration of obesity and overweight and risk of type 2 diabetes among US women. *Obesity*, 22(10):2267–2273.

[Huxley et al., 2010] Huxley, R., Mendis, S., Zheleznyakov, E., Reddy, S., and Chan, J. (2010). Body mass index, waist circumference and waist:hip ratio as predictors of cardiovascular risk: a review of the literature. *European Journal of Clinical Nutrition*, 64(1):16–22.

[Jensen et al., 2013] Jensen, M. D., Ryan, D. H., Apovian, C. M., Ard, J. D., Comuzzie, A. G., Donato, K. A., Hu, F. B., Hubbard, V. S., Jakicic, J. M., Kushner, R. F., et al. (2013). 2013 aha/acc/tos guideline for the management of overweight and obesity in adults. *Circulation*, pages 01–cir.

[Joffe, 2016] Joffe, A. (2016). Adolescent obesity and functional limitations in young adulthood. *NEJM Journal Watch: Pediatrics & Adolescent Medicine*.

[Johnson and Wardle, 2011] Johnson, F. and Wardle, J. (2011). Socio-economic status and obesity: epidemiology and explanations. *Psychologist*, 24(12):893–893.

[Johnson and Johnson, 2015] Johnson, J. A. and Johnson, A. M. (2015). Urban-rural differences in childhood and adolescent obesity in the United States: A systematic review and meta-analysis. *Childhood Obesity*, 11(3):233–241.

[Jull et al., 2011] Jull, A., Lawes, C. M. M., Eyles, H., Maddison, R., Gorton, D., Arcus, K., Chee, N., Taylor, B., and Mann, J. (2011). Clinical guidelines for weight management in new zealand adults, children and young people. *Journal of Primary Health Care*, 3(1):66–71.

[Juonala et al., 2011] Juonala, M., Magnussen, C. G., Berenson, G. S., Venn, A., Burns, T. L., Sabin, M. A., Srinivasan, S. R., Daniels, S. R., Davis, P. H., Chen, W., Sun, C., Cheung, M., Viikari, J. S. A., Dwyer, T., and Raitakari, O. T. (2011). Childhood adiposity, adult adiposity, and cardiovascular risk factors. *New England Journal of Medicine*, 365(20):1876–1885.

[Kaiser Family Foundation, 2016] Kaiser Family Foundation (2016). Population distribution by race/ethnicity. `http://kff.org/other/state-indicator/distribution-by-raceethnicity/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D`. Accessed: 5 June 2017.

[Kearney, 2010] Kearney, J. (2010). Food consumption trends and drivers. *Philosophical Transactions: biological Sciences*, 365(1554):2793–2807.

[Kelly and Swinburn, 2015] Kelly, S. and Swinburn, B. (2015). Childhood obesity in New Zealand. *The New Zealand medical journal*, 128(1417):6–7.

[Kennedy et al., 2015] Kennedy, S., Kidd, M. P., McDonald, J. T., and Biddle, N. (2015). The healthy immigrant effect: patterns and evidence from four countries. *Journal of International Migration and Integration*, 16(2):317–332.

[Kovacs et al., 2014] Kovacs, E., Siani, A., Konstabel, K., Hadjigeorgiou, C., de Bourdeaudhuij, I., Eiben, G., Lissner, L., Gwozdz, W., Reisch, L., Pala, V., Moreno, L. A., Pigeot, I., Pohlabeln, H., Ahrens, W., and Molnar, D. (2014). Adherence to the obesity-related lifestyle intervention targets in the IDEFICS study. *International Journal of Obesity*, 38(S2):S144–S151.

[Lai et al., 2014] Lai, C., Sun, D., Cen, R., Wang, J., Li, S., Fernandez-Alonso, C., Chen, W., Srinivasan, S. R., and Berenson, G. S. (2014). Impact of long-term burden of excessive adiposity and elevated blood pressure from childhood on adulthood left ventricular remodeling patterns: the Bogalusa Heart Study. *Journal of the American College of Cardiology*, 64(15):1580–1587.

[Le et al., 2016] Le, H., Engler-Stringer, R., and Muhajarine, N. (2016). Walkable home neighbourhood food environment and children's overweight and obesity: proximity, density or price? *Canadian Journal of Public Health*, 107(1):ES42–ES47.

[Ledoux et al., 2011] Ledoux, T. A., Hingle, M. D., and Baranowski, T. (2011). Relationship of fruit and vegetable intake with adiposity: a systematic review. *Obesity Reviews*, 12(501):e143–e150.

[Li et al., 2008] Li, G., Li, H., Li, H., Zhang, P., Zhang, J., Zhang, B., Wang, J., Gregg, E. W., Yang, W., Gong, Q., Jiang, Y., An, Y., Shuai, Y., Thompson, T. J., Gerzoff, R. B., Roglic, G., Hu, Y., and Bennett, P. H. (2008). The long-term effect of lifestyle interventions to prevent diabetes in the China Da Qing Diabetes Prevention Study: a 20-year follow-up study. *The Lancet*, 371(9626):1783–1789.

[Li et al., 2014] Li, X., Memarian, E., Sundquist, J., Zöller, B., and Sundquist, K. (2014). Neighbourhood deprivation, individual-level familial and socio-demographic factors and diagnosed childhood obesity: a nationwide multilevel study from Sweden. *Obesity facts*, 7(4):253–263.

[Luke and Cooper, 2013] Luke, A. and Cooper, R. S. (2013). Physical activity does not influence obesity risk: time to clarify the public health message. *Internatonal Journal of Epidemiology*, 42(6):1831–1836.

[Luke et al., 2011] Luke, A., Dugas, L. R., Durazo-Arvizu, R. A., Cao, G., and Cooper, R. S. (2011). Assessing physical activity and its relationship to cardiovascular risk factors: Nhanes 2003-2006. *BMC Public Health*, 11(1):387–387.

[Lumley, 2004] Lumley, T. (2004). Analysis of complex survey samples. *Journal of Statistical Software*, 9(1):1–19. R package verson 2.2.

[Lumley, 2016] Lumley, T. (2016). Survey: analysis of complex survey samples. R package version 3.31-5.

[MacKay et al., 2013] MacKay, D. F., Gray, L., and Pell, J. P. (2013). Impact of smoking and smoking cessation on overweight and obesity: Scotland-wide, cross-sectional study on 40,036 participants. *BMC Public Health*, 13(1):348–348.

[Maersk et al., 2012] Maersk, M., Belza, A., Holst, J. J., Fenger-Grøn, M., Pedersen, S. B., Astrup, A., and Richelsen, B. (2012). Satiety scores and satiety hormone response after sucrose-sweetened soft drink compared with isocaloric semi-skimmed milk and with non-caloric soft drink: a controlled trial. *European Journal of Clinical Nutrition*, 66(4):523–529.

[Maier et al., 2013] Maier, W., Holle, R., Hunger, M., Peters, A., Meisinger, C., Greiser, K., Kluttig, A., Völzke, H., Schipf, S., Moebus, S., et al. (2013). The impact of regional deprivation and individual socio-economic status on the prevalence of type 2 diabetes in germany. a pooled analysis of five population-based studies. *Diabetic Medicine*, 30(3).

[Malik et al., 2013a] Malik, V. S., Pan, A., Willett, W. C., and Hu, F. B. (2013a). Sugar-sweetened beverages and weight gain in children and adults: A systematic review and meta-analysis. *American Journal of Clinical Nutrition*, 98(4):1084–1102.

[Malik et al., 2010] Malik, V. S., Popkin, B. M., Bray, G. A., Després, J.-P., Willett, W. C., and Hu, F. B. (2010). Sugar-sweetened beverages and risk of metabolic syndrome and type 2 diabetes. *Diabetes Care*, 33(11):2477–2483.

[Malik et al., 2013b] Malik, V. S., Willett, W. C., and Hu, F. B. (2013b). Global obesity: trends, risk factors and policy implications. *Nature Reviews Endocrinology*, 9(1):13–27.

[Matheson et al., 2008] Matheson, F. I., Moineddin, R., and Glazier, R. H. (2008). The weight of place: A multilevel analysis of gender, neighborhood material deprivation, and body mass index among canadian adults. *Social Science & Medicine*, 66(3):675–690.

[Mattes and Foster, 2014] Mattes, R. and Foster, G. D. (2014). Food environment and obesity. *Obesity*, 22(12):2459–2461.

[McCormack and Shiell, 2011] McCormack, G. R. and Shiell, A. (2011). In search of causality: A systematic review of the relationship between the built environment and physical activity among adults. *International Journal of Behavioral Nutrition and Physical Activity*, 8(1):125–125.

[Mei et al., 2012] Mei, H., Chen, W., Jiang, F., He, J., Srinivasan, S., Smith, E. N., Schork, N., Murray, S., and Berenson, G. S. (2012). Longitudinal replication studies of GWAS risk SNPs influencing body mass index over the course of childhood and adulthood. *PLoS ONE*, 7(2):e31470.

[Min et al., 2013] Min, J., Chiu, D. T., and Wang, Y. (2013). Variation in the heritability of body mass index based on diverse twin studies: a systematic review. *Obesity Reviews*, 14(11):871–882.

[Ministry of Health, 2009] Ministry of Health (2009). *Report on New Zealand cost-of-illness studies on long-term conditions*. Ministry of Health, Wellington.

[Ministry of Health, 2015a] Ministry of Health (2015a). *Annual update of key results 2014/15: New Zealand health survey*. Ministry of Health, Wellington.

[Ministry of Health, 2015b] Ministry of Health (2015b). *Methodology report for the 2014/15 New Zealand Health Survey*. Ministry of Health, Wellington.

[Ministry of Health, 2016] Ministry of Health (2016). *Annual update of key results 2015/16: New Zealand health survey*. Ministry of Health, Wellington.

[Morrissey et al., 2014] Morrissey, T. W., Jacknowitz, A., and Vinopal, K. (2014). Local food prices and their associations with children's weight and food security. *Pediatrics*, 133(3):422–430.

[Nau et al., 2015] Nau, C., Schwartz, B. S., BandeenâĂŘRoche, K., Liu, A., Pollak, J., Hirsch, A., BaileyâĂŘ-Davis, L., and Glass, T. A. (2015). Community socioeconomic deprivation and obesity trajectories in children using electronic health records. *Obesity*, 23(1):207–212.

[Neff et al., 2013] Neff, K. J., Olbers, T., and le Roux, C. W. (2013). Bariatric surgery: the challenges with candidate selection, individualizing treatment and clinical outcomes. *BMC Medicine*, 11(1):8.

[Newton et al., 2017] Newton, S., Braithwaite, D., and Akinyemiju, T. (2017). Socio-economic status over the life course and obesity: systematic review and meta-analysis. *PLoS ONE*, 12(5).

[Ni Mhurchu and Ogra, 2007] Ni Mhurchu, C. and Ogra, S. (2007). The price of healthy eating: cost and nutrient value of selected regular and healthier supermarket foods in New Zealand. *The New Zealand Medical Journal*, 120(1248):U2388.

[Norredam et al., 2014] Norredam, M., Agyemang, C., Hansen, H., Oluf, K., Petersen, J. H., Byberg, S., Krasnik, A., and Kunst, A. E. (2014). Duration of residence and disease occurrence among refugees and family reunited immigrants: test of the "healthy migrant" effect hypothesis. *Tropical Medicine & International Health*, 19(8):958–967.

[OECD, 2017] OECD (2017). Obesity update 2017. `http://www.oecd.org/els/health-systems/Obesity-Update-2017.pdf`. Accessed: 11 July 2017.

[Office for National Statistics, 2011] Office for National Statistics (2011). Ethnicity and national identity in England and Wales: 2011. `https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/articles/ethnicityandnationalidentityinenglandandwales/2012-12-11`. Accessed: 22 June 2017.

[Ogden et al., 2013] Ogden, C. L., Carroll, M. D., Kit, B. K., and Flegal, K. M. (2013). Prevalence of obesity among adults: United States, 2011-2012. *NCHS data brief*, (131):1.

[O'Rahilly and Farooqi, 2008] O'Rahilly, S. and Farooqi, I. S. (2008). Human obesity: a heritable neurobehavioral disorder that is highly sensitive to environmental conditions. *Diabetes*, 57(11):2905–2910.

[Padwal et al., 2003] Padwal, R. S., Rucker, D., Li, S. K., Curioni, C., and Lau, D. C. W. (2003). Long-term pharmacotherapy for obesity and overweight. *The Cochrane Library*.

[Pedrosa et al., 2011] Pedrosa, C., Correia, F., Seabra, D., Oliveira, B. M., Simoes-Pereira, C., and Vaz-de-Almeida, M. D. (2011). Prevalence of overweight and obesity among 7-9-year-old children in Aveiro, Portugal: comparison between IOTF and CDC references. *Public Health Nutrition*, 14(01):14–19.

[Pereira et al., 2005] Pereira, M. A., Kartashov, A. I., Ebbeling, C. B., Van Horn, L., Slattery, M. L., Jacobs, D. R., and Ludwig, D. S. (2005). Fast-food habits, weight gain, and insulin resistance (the CARDIA study): 15-year prospective analysis. *The lancet*, 365(9453):36–42.

[Powell et al., 2009] Powell, L. M., Zhao, Z., and Wang, Y. (2009). Food prices and fruit and vegetable consumption among young American adults. *Health and Place*, 15(4):1064–1070.

[Rigotti and Clair, 2013] Rigotti, N. A. and Clair, C. (2013). Managing tobacco use: the neglected cardiovascular disease risk factor. *European Heart Journal*, 34(42):3259–3267.

[Rom et al., 2015] Rom, O., Reznick, A. Z., Keidar, Z., Karkabi, K., and Aizenbud, D. (2015). Smoking cessation-related weight gain-beneficial effects on muscle mass, strength and bone health. *Addiction*, 110(2):326–335.

[Romero-Corral et al., 2006] Romero-Corral, A., Montori, V. M., Somers, V. K., Korinek, J., Thomas, R. J., Allison, T. G., Mookadam, F., and Lopez-Jimenez, F. (2006). Association of bodyweight with total mortality and with cardiovascular events in coronary artery disease: a systematic review of cohort studies. *The Lancet*, 368(9536):666–678.

[Rosenheck, 2008] Rosenheck, R. (2008). Fast food consumption and increased caloric intake: a systematic review of a trajectory towards weight gain and obesity risk. *Obesity Reviews*, 9(6):535–547.

[Rush et al., 2009] Rush, E. C., Freitas, I., and Plank, L. D. (2009). Body size, body composition and fat distribution: comparative analysis of European, Maori, Pacific Island and Asian Indian adults. *British Journal of Nutrition*, 102(4):632–641.

[Shaw et al., 2006] Shaw, K. A., Gennat, H. C., O'Rourke, P., and Mar, C. D. (2006). Exercise for overweight or obesity. *The Cochrane Library*.

[Sheperd, 2005] Sheperd, T. M. (2005). Bariatric surgery in context. *Journal of Family Practice*, 54(2):3–9.

[Siahpush et al., 2014] Siahpush, M., Singh, G., Tibbits, M., Pinard, C., Shaikh, R., and Yaroch, A. (2014). It is better to be a fat ex-smoker than a thin smoker: findings from the 1997-2004 national health interview survey-national death index linkage study. *Tobacco Control*, 23(5):395–402.

[Singh et al., 2015] Singh, G. K., Siahpush, M., Azuine, R. E., and Williams, S. D. (2015). Increasing area deprivation and socioeconomic inequalities in heart disease, stroke, and cardiovascular disease mortality among working age populations, united states, 1969-2011. *International Journal of MCH and AIDS*, 3(2):119–133.

[Sjöström et al., 2012] Sjöström, L., Peltonen, M., Jacobson, P., Sjöström, C. D., Karason, K., Wedel, H., Ahlin, S., Anveden, Å., Bengtsson, C., Bergmark, G., et al. (2012). Bariatric surgery and long-term cardiovascular events. *JAMA*, 307(1):56–65.

[Smith et al., 2010] Smith, C., Parnell, W., and Brown, R. (2010). *Family food environment: barriers to acquiring affordable and nutritious food in New Zealand households*. Families Commission.

[Smith et al., 2013] Smith, C., Parnell, W. R., Brown, R. C., and Gray, A. R. (2013). Balancing the diet and the budget: food purchasing practices of food-insecure families in New Zealand. *Nutrition & Dietetics*, 70(4):278–285.

[Stafford et al., 2010] Stafford, M., Brunner, E. J., Head, J., and Ross, N. A. (2010). Deprivation and the development of obesity. *American Journal of Preventive Medicine*, 39(2):130–139.

[Statistics New Zealand, 2017] Statistics New Zealand (2017). Ethnic group (total responses) by age group and sex, for the census usually resident population count, 2001, 2006, and 2013 censuses (RC, TA). `http://nzdotstat.stats.govt.nz/wbos/Index.aspx?DataSetCode=TABLECODE8021`. Accessed: 25 May 2017.

[Stenberg, 2013] Stenberg, A. (2013). Interpreting estimates of heritability–a note on the twin decomposition. *Economics & Human Biology*, 11(2):201–205.

[Stommel and Schoenborn, 2010] Stommel, M. and Schoenborn, C. A. (2010). Variations in bmi and prevalence of health risks in diverse racial and ethnic populations. *Obesity*, 18(9):1821–1826.

[Sturm and An, 2014] Sturm, R. and An, R. (2014). Obesity and economic environments. *CA: A Cancer Journal for Clinicians*, 64(5):337–350.

[Sturm and Datar, 2011] Sturm, R. and Datar, A. (2011). Regional price differences and food consumption frequency among elementary school children. *Public Health*, 125(3):136–141.

[Swinburn et al., 2014] Swinburn, B., Dominick, C., and Vandevijvere, S. (2014). *Benchmarking food environments: experts′ assessments of policy gaps and priorities for the New Zealand Government*. The University of Auckland, Auckland.

[Tian et al., 2014] Tian, X., Zhao, G., Li, Y., Wang, L., and Shi, Y. (2014). Overweight and obesity difference of Chinese population between different urbanization levels: overweight and obesity between urbanization levels. *The Journal of Rural Health*, 30(1):101–112.

[Vandevijvere et al., 2015] Vandevijvere, S., Chow, C. C., Hall, K. D., Umali, E., and Swinburn, B. A. (2015). Increased food energy supply as a major driver of the obesity epidemic: a global analysis. *Bulletin of The World Organization*, 93(7):446–456.

[Vang et al., 2015] Vang, Z., Sigouin, J., Flenon, A., and Gagnon, A. (2015). The healthy immigrant effect in canada: a systematic review. *Population Change and Lifecourse Strategic Knowledge Cluster Discussion Paper Series*, 3(1):4.

[Vartanian et al., 2007] Vartanian, L. R., Schwartz, M. B., and Brownell, K. D. (2007). Effects of soft drink consumption on nutrition and health: a systematic review and meta-analysis. *American Journal of Public Health*, 97(4):667–675.

[Vazquez et al., 2007] Vazquez, G., Duval, S., Jacobs, D. R., and Silventoinen, K. (2007). Comparison of body mass index, waist circumference, and waist/hip ratio in predicting incident diabetes: a meta-analysis. *Epidemiologic reviews*, 29(1):115–128.

[Visscher et al., 2012] Visscher, P. M., Brown, M. A., McCarthy, M. I., and Yang, J. (2012). Five years of GWAS discovery. *American Journal of Human Genetics*, 90(1):7–24.

[Wadden et al., 2014] Wadden, T. A., Bantle, J. P., Blackburn, G. L., Bolin, P., Brancati, F. L., Bray, G. A., Clark, J. M., Coday, M., Dutton, G. R., Egan, C., Evans, M., Foreyt, J. P., Sengardi, S. G., Gregg, E. W., Hazuda, H. P., Hill, J. O., Horton, E. S., Hubbard, V. S., Jakicic, J. M., Jeffery, R. W., Johnson, K. C., Kahn, S. E., Kitabchi, A. E., Knowler, W. C., Lewis, C. E., Maschak-Carey, B. J., Montez, M. G., Montgomery, B., Nathan, D. M., Nelson, J., Patricio, J., Peters, A., Pi-Sunyer, F. X., Pownall, H., Rickman, A. D., Vitolins, M., Walkup, M. P., West, D. S., Williamson, D., Wing, R. R., Wyatt, H., Yanovski, S. Z., and Look Ahead Research Group (2014). Eight-year weight losses with an intensive lifestyle intervention: The look AHEAD study. *Obesity*, 22(1):5–13.

[Wang et al., 2010] Wang, J., Williams, M., Rush, E., Crook, N., Forouhi, N. G., and Simmons, D. (2010). Mapping the availability and accessibility of healthy food in rural and urban New Zealand — Te Wai o Rona: Diabetes Prevention Strategy. *Public Health Nutrition*, 13(7):1049–1055.

[Weiss and Kaufman, 2008] Weiss, R. and Kaufman, F. R. (2008). Metabolic complications of childhood obesity: identifying and mitigating the risk. *Diabetes care*, 31(2):S310–316.

[Wing et al., 2010] Wing, R. R., Bahnson, J. L., Bray, G. A., Clark, J. M., Coday, M., Egan, C., Espeland, M. A., Foreyt, J. P., Gregg, E. W., Goldman, V., Haffner, S. M., Hazuda, H., Hill, J. O., Horton, E. S., Hubbard, V. S., Jakicic, J., Jeffery, R. W., Johnson, K. C., Kahn, S., Killean, T., Kitabchi, A. E., Lewis, C. E., Manus, C., Maschak-Carey, B. J., Michaels, S., Montez, M., Montgomery, B., Nathan, D. M., Patricio, J., Peters, A., Pi-Sunyer, X., Pownall, H., Reboussin, D., Rejeski, W. J., Rubin, R., Safford, M., Skarphol, T., Van Dorsten, B., Wadden, T. A., Wagenknecht, L., Wesche-Thobaben, J., West, D. S., Williamson, D., Yanovski, S. Z., and Group, L. A. R. (2010). Long-term effects of a lifestyle intervention on weight and cardiovascular risk factors in individuals with type 2 diabetes mellitus: four-year results of the look AHEAD trial. *Archives of Internal Medicine*, 170(17):1566–1575.

[Woolf and Nestle, 2008] Woolf, S. H. and Nestle, M. (2008). Do dietary guidelines explain the obesity epidemic? *American Journal of Preventive Medicine*, 34(3):263–265.

[World Health Organization, 2015] World Health Organization (2015). Obesity and overweight. `http://www.who.int/mediacentre/factsheets/fs311/en/`. Accessed: 22 April 2017.

[World Obesity Federation,] World Obesity Federation. Extended international (iotf) body mass index cut-offs for thinness, overweight and obesity in children. `http://www.worldobesity.org/data/cut-points-used/newchildcutoffs/`. Accessed: 13 July 2017.

[Wu et al., 2015] Wu, S. Q., Ding, Y. Y., Wu, F. Q., Li, R. S., Hu, Y., Hou, J., and Mao, P. Y. (2015). Socio-economic position as an intervention against overweight and obesity in children: a systematic review and meta-analysis. *Scientific Reports*, 5:11354.

[Xue et al., 2016] Xue, H., Wu, Y., Wang, X., and Wang, Y. (2016). Time trends in fast food consumption and its association with obesity among children in China. *PLoS ONE*, 11(3):e0151141.

[Zhang et al., 2017] Zhang, J., Wang, D., Eldridge, A. L., Huang, F., Ouyang, Y., Wang, H., and Zhang, B. (2017). Urban-rural disparities in energy intake and contribution of fat and animal source foods in Chinese children aged 4-17 years. *Nutrients*, 9(5):526.

A R Codes

Listing A.1: Cleaning The NZHSs Data

```r

#### DATA CLEANING – New Zealand Health Surveys

## Load NZHS data from 2002/03 to 2014/15 into the system. Subsetting is done in the Microsoft Excel
    2016.
# Subsetting done in the Microsoft Excel is not essential. It only helps to reduce to the time to
    load the data.
HS02 <- read.csv("D:/UC/Thesis/Data/NZHS/HS02-subset.csv", na.strings=c("","NA"))
HS06A <- read.csv("D:/UC/Thesis/Data/NZHS/HS06A-subset.csv", na.strings=c("","NA"))
HS06C <- read.csv("D:/UC/Thesis/Data/NZHS/HS06C-subset.csv", na.strings=c("","NA"))
HS11A <- read.csv("D:/UC/Thesis/Data/NZHS/HS11A-subset.csv",na.strings=c("","NA"))
HS11C <- read.csv("D:/UC/Thesis/Data/NZHS/HS11C-subset.csv",na.strings=c("","NA"))
HS12A <- read.csv("D:/UC/Thesis/Data/NZHS/HS12A-subset.csv", na.strings=c("","NA"))
HS12C <- read.csv("D:/UC/Thesis/Data/NZHS/HS12C-subset.csv", na.strings=c("","NA"))
HS13A <- read.csv("D:/UC/Thesis/Data/NZHS/HS13A-subset.csv", na.strings=c("","NA"))
HS13C <- read.csv("D:/UC/Thesis/Data/NZHS/HS13C-subset.csv", na.strings=c("","NA"))
HS14A <- read.csv("D:/UC/Thesis/Data/NZHS/HS14A-subset.csv", na.strings=c("","NA"))
HS14C <- read.csv("D:/UC/Thesis/Data/NZHS/HS14C-subset.csv", na.strings=c("","NA"))


#### Install Packages ####

# package to do crosstable
install.packages("gmodels")

#package to draw chart, graph, etc.
install.packages("ggplot2")

# package for text dodge in charts and graphs
install.packages("ggrepel")

# package for a colour pallete
install.packages("RColorBrewer")

```

```r
# package for creating mosaic plot
install.packages("ggmosaic")

# package for transforming and re-arranging data
install.packages("dplyr")
install.packages("tidyr")

# package for wrapping long text
install.packages("stringr")

# package to arrange plot
install.packages("gridExtra")

# package to adjust plots' size in gridExtra
install.packages('cowplot')

# function for multiple crosstabs
source("http://pcwww.liv.ac.uk/~william/R/crosstab.r")

# package for confidence interfal and effect size calculator
install.packages("MBESS")

# package for creating a survey desing object
install.packages("survey")
install.packages("segmented")

# package for reading a sas7bdat format
install.packages("sas7bdat")

# package for cutting strings
install.packages("stringi")

# package for calculating kurtosis and skewness
install.packages("moments")

# package for evaluating environment inside tidyr and dyplr functions
install.packages("lazyeval")

# package for drawing interactive charts
install.packages("plotly")
```

```r


#### Load the packages ####
library(gmodels)
library(ggplot2)
library(ggmosaic)
library(ggrepel)
library(RColorBrewer)
library(stringi)
library(coin)
library(scales)
library(moments)
library(tidyr)
library(plyr)
library(dplyr)
library(stringr)
library(gridExtra)
library(cowplot)
library(lazyeval)
library(MBESS)
library(plotly)
library(survey)
library(segmented)


#### Function Sets ####
## Function set 1 (proportion table and several statistic tests)

# Colourblind Palete
GreyPalette <- c("#d3d3d3","#bdbdbd","#a8a8a8","#939393","#7e7e7e","#696969","#545454","#3f3f3f","#2a2a2a","#151515","#000000")
cbbPalette <- c("#000000", "#E69F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2", "#D55E00", "#CC79A7")


# Cramer's v function
cv.test <- function(x,y) {
  # Cramer's V function
  CV = sqrt(chisq.test(x, y, correct=FALSE)$statistic /
    (length(x) * (min(length(unique(x)),length(unique(y))) - 1)))
```

```r
  return(as.numeric(CV))
}


# A function for non-linear test
non.lin <- function(x,y){
  # non-linear association chi square
    X2 = (chisq.test(x, y, correct=FALSE)$statistic) - (statistic(lbl_test(table(x,y)))^2)
    p = 1-pchisq(as.numeric(X2),(min(length(unique(x)),length(unique(y))) - 1))
  print.noquote("X2 non linear & p value")
  return(cbind(as.numeric(X2),as.numeric(p)))
}


# A function for creating proportion table by groups
prop_frame <- function(data, var1) {
  # a function to show number and proportion of grouped variables, use quote on var1

  df <- data %>%
    group_by_(var1) %>%
     filter_(interp(~ !is.na(var), var = as.name(var1))) %>%
    summarise(n=n()) %>%
    mutate(prop=n/sum(n))

  df$perc = round(df$prop*100, 2)
  df$cop = sprintf("%s (%s)", df$n, df$perc)

  return(df)
}



## Function set 2 (normal plot, charts, graphs)

# A function for displaying histogram and qq plot
normal_plot <- function(data, var1){
  # histogram and qq plot function. Use quote for var1 (eg. var1 = "bmiscale")

  vector <- as.numeric(unlist(data[var1]))

```

```
  histogram <- ggplot(data, aes_string(var1)) +
    geom_histogram(aes(y=..density..), bins= 50, alpha= 0.5, na.rm= TRUE) +
    geom_density(col=2, na.rm= TRUE) +
    scale_x_continuous(limits=c(18,95), breaks=seq(10,100,10)) +
    scale_y_continuous()


  y <- quantile(vector[!is.na(vector)], c(0.25, 0.75))
  x <- qnorm(c(0.25, 0.75))
  slope <- diff(y)/diff(x)
  int <- y[1L] - (slope*x[1L])

  qq <- ggplot(data, aes_string(sample=var1)) +
    stat_qq() +
    geom_abline(slope= slope, intercept= int, color="red") +
    scale_y_continuous()

  grid <- grid.arrange(histogram, qq, ncol=2)

  return(grid)
}



# function for box plot + showing median
box_plot_median <- function(data, var_x, var_y, title_x, title_y, year, age=">=18 yo", txt.size = 3,
                            t.size = 15, ax.size = 10, x.size = 10, y.size = 10,
                            min=0, max=100, bins = 10, wide=0.8, gap= 2){
  # box plot for numerical scale by categorical group, use quote for var_x ~ year
  # year is just a string of text for the title

  complete_data <- data %>%
    select_(var_x, var_y) %>%
    filter_(interp(~!is.na(var), var= as.name(var_x)))

  colnames(complete_data) <- c("var1","var2")

  frame <- complete_data %>%
    group_by(var1) %>%
    summarise_each(funs(median), var2)
```

```
  colnames(frame)[2] <- "number"

  plot <- ggplot(complete_data, aes(x=var1, y=var2, na.rm=TRUE), environment=environment()) +
    geom_boxplot(width=wide) +
    scale_y_continuous(breaks=seq(0,100,bins), limits=c(min,max)) +
    labs(title=sprintf("%s by %s \n(%s, age)", title_y, title_x, year),
         x= sprintf("%s", title_x), y= sprintf("%s", title_y)) +
    theme(axis.title = element_text(size= ax.size, face="bold"),
          plot.title= element_text(size= t.size, face="bold", hjust=0.5),
          axis.text.x = element_text(size = x.size),
          axis.text.y = element_text(size = y.size)) +
    geom_text(data=frame, aes(label=round(frame$number,1)), y=frame$number+gap, size= txt.size) +
    scale_x_discrete(label= function(x) str_wrap(x, width=10))

  return(plot)
}


# box chart without title
box_notitle <- function(data, var_x, var_y, title_x, title_y, txt.size = 3,
                        t.size = 10, ax.size = 6, x.size = 8, y.size = 8,
                        min=0, max=100, bins = 10, wide=0.8, gap= 3){
  # box plot for numerical scale by categorical group, use quote for var_x ~ year
  # year is just a string of text for the title

  complete_data <- data %>%
    select_(var_x, var_y) %>%
    filter_(interp(~!is.na(var), var= as.name(var_x)))

  colnames(complete_data) <- c("var1","var2")

  frame <- complete_data %>%
    group_by(var1) %>%
    summarise_each(funs(median), var2)

  colnames(frame)[2] <- "number"

  plot <- ggplot(complete_data, aes(x=var1, y=var2, na.rm=TRUE), environment=environment()) +
    geom_boxplot(width=wide) +
```

```
    scale_y_continuous(breaks=seq(0,100,bins), limits=c(min,max)) +
    labs(title=sprintf("%s by %s", title_y, title_x),
         x= sprintf("%s", title_x), y= sprintf("%s", title_y)) +
    theme(axis.title = element_text(size= ax.size, face="bold"),
          plot.title= element_text(size= t.size, face="bold", hjust=0.5),
          axis.text.x = element_text(size = x.size),
          axis.text.y = element_text(size = y.size)) +
    geom_text(data=frame, aes(label=round(frame$number,1)), y=frame$number+gap, size= txt.size) +
    scale_x_discrete(label= function(x) str_wrap(x, width=10))

  return(plot)
}


# function for stacked plot + showing percentage
stacked_plot <- function(data, var_x, var_y, title1, title2, year, age = ">=18 yo", ax.size = 10,
                         t.size = 15, l.size = 10, x.size = 10, y.size = 10, txt.size = 3){
  # stacked bar chart, use quote for var_x ~ age
  # year is just a string of text

  data_frame <- data %>%
    filter_(interp(~ !is.na(var), var = as.name(var_x)),
            interp(~ !is.na(var), var = as.name(var_y))) %>%
    group_by_(var_x, var_y) %>%
    summarise(count=n()) %>%
    mutate(percent = count/sum(count),
           per = count/sum(count)*100,
           pos = (cumsum(percent) - 0.5*percent)) # calculating position

  # labels
  data_frame$label_percent <- paste0(sprintf("%.1f", data_frame$per),"%")

  colnames(data_frame)[1:2] <- c("var_x", "var_y")

  localenv <- environment()

  plot <- ggplot(data_frame, aes(x= var_x, y=percent, fill=factor(var_y, levels=rev(var_y))),
      environment= localenv) +
    geom_bar(stat="identity", width=0.8) +
    labs(title= sprintf("%s by %s \n(%s, %s)", title1, title2, year, age),
```

```
        x= sprintf("%s", title2), y= "Density") +
    scale_fill_discrete(name = sprintf("%s", title1)) +
    theme(axis.title = element_text(size = ax.size, face="bold"),
          plot.title = element_text(size = t.size, face="bold", hjust=0.5),
          legend.title = element_text(size = l.size, face="bold"),
          axis.text.x = element_text(size = x.size),
          axis.text.y = element_text(size = y.size)) +
    geom_label(aes(y=data_frame$pos, label=data_frame$label_percent), size=txt.size, fill="White") +
    scale_x_discrete(label= function(x) str_wrap(x, width=8))

  return(plot)
}

# stacked bar chart without title
stacked_notitle <- function(data, var_x, var_y, title1, title2, ax.size = 6,
                            t.size = 10, l.size = 8, x.size = 8, y.size = 8, lt.size = 8, txt.size = 3)
                            {
  # stacked bar chart, use quote for var_x ~ age
  # year is just a string of text

  data_frame <- data %>%
    filter_(interp(~ !is.na(var), var = as.name(var_x)),
            interp(~ !is.na(var), var = as.name(var_y))) %>%
    group_by_(var_x, var_y) %>%
    summarise(count=n()) %>%
    mutate(percent = count/sum(count),
           per = count/sum(count)*100,
           pos = (cumsum(percent) - 0.5*percent)) # calculating position

  # labels
  data_frame$label_percent <- paste0(sprintf("%.1f", data_frame$per),"%")

  colnames(data_frame)[1:2] <- c("var_x", "var_y")

  localenv <- environment()

  plot <- ggplot(data_frame, aes(x= var_x, y=percent, fill=factor(var_y, levels=rev(var_y))),
      environment= localenv) +
    geom_bar(stat="identity", width=0.8) +
    labs(title= sprintf("%s by %s", title1, title2),
```

```
        x= sprintf("%s", title2), y= "Density") +
  scale_fill_discrete(name = sprintf("%s", title1)) +
  theme(axis.title = element_text(size = ax.size, face="bold"),
        plot.title = element_text(size = t.size, face="bold", hjust=0.5),
        legend.title = element_text(size = l.size, face="bold"),
        axis.text.x = element_text(size = x.size),
        axis.text.y = element_text(size = y.size),
        legend.text=element_text(size = lt.size)) +
  geom_text(aes(y=data_frame$pos, label=data_frame$label_percent), size=txt.size) +
  scale_x_discrete(label= function(x) str_wrap(x, width=8))

 return(plot)
}


# stacked bar chart for ethnicity
stacked_eth <- function(data, var_x, var_y, title1, title2, ax.size = 6,
                        t.size = 10, l.size = 8, x.size = 8, y.size = 8, lt.size = 8, txt.size = 2)
                            {
 # stacked bar chart, use quote for var_x ~ age
 # year is just a string of text

 data_frame <- data %>%
   filter_(interp(~ !is.na(var), var = as.name(var_x)),
           interp(~ !is.na(var), var = as.name(var_y))) %>%
   group_by_(var_x, var_y) %>%
   summarise(count=n()) %>%
   mutate(percent = count/sum(count),
          per = count/sum(count)*100,
          pos = (cumsum(percent) - 0.5*percent)) # calculating position

 # labels
 data_frame$label_percent <- paste0(sprintf("%.1f", data_frame$per),"%")

 colnames(data_frame)[1:2] <- c("var_x", "var_y")

 localenv <- environment()

 plot <- ggplot(data_frame, aes(x= var_x, y=percent, fill=factor(var_y, levels=rev(var_y))),
    environment= localenv) +
```

```r
    geom_bar(stat="identity", width=0.8) +
    labs(title= sprintf("%s by %s", title1, title2),
         x= sprintf("%s", title2), y= "Density") +
    scale_fill_discrete(name = sprintf("%s", title1)) +
    theme(axis.title = element_text(size = ax.size, face="bold"),
          plot.title = element_text(size = t.size, face="bold", hjust=0.5),
          legend.title = element_text(size = l.size, face="bold"),
          axis.text.x = element_text(size = x.size),
          axis.text.y = element_text(size = y.size),
          legend.text=element_text(size = lt.size)) +
    geom_text(aes(y=data_frame$pos, label=data_frame$label_percent), size=txt.size, position = position_dodge(width=1)) +
    scale_x_discrete(label= function(x) str_wrap(x, width=8))

  return(plot)
}



# stacked bar chart without label
stacked_nolabel <- function(data, var_x, var_y, title1, title2, year){
  # stacked bar chart without label, use quote for var_x ~ year
  # year is just a string of text

  data_frame <- data %>%
    filter_(interp(~ !is.na(var), var = as.name(var_x)), interp(~ !is.na(var), var = as.name(var_y))) %>%
    group_by_(var_x, var_y) %>%
    summarise(count=n()) %>%
    mutate(percent = count/sum(count),
           per = count/sum(count)*100)

  colnames(data_frame)[1:2] <- c("var_x", "var_y")

  localenv <- environment()

  plot <- ggplot(data_frame, aes(x= var_x, y=percent, fill=var_y), environment= localenv) +
    geom_bar(stat="identity", width=0.8) +
    labs(title= sprintf("%s by %s \n(%s, >=18 yo)", title2, title1, year),
         x= sprintf("%s", title1), y= "Density") +
```

```r
    scale_fill_discrete(name = sprintf("%s", title2)) +
    theme(axis.title = element_text(face="bold"), plot.title= element_text(face="bold", hjust=0.5),
          legend.title=element_text(face="bold")) +
    scale_x_discrete(label= function(x) str_wrap(x, width=10))

  return(plot)
}



## Function set 3 (custom statistic tables + effect sizes)

# function for anova table + effect size confidence interval
anova_table <- function(df1, df2){
  # neat anova table with effect size CI
  df_between <- anova(aov(df1 ~ df2))[1,1]
  df_resid <- anova(aov(df1 ~ df2))[2,1]
  F_value <- anova(aov(df1 ~ df2))[1,4]
  p_value <- anova(aov(df1 ~ df2))[1,5]
  eta_squared <- anova(aov(df1 ~ df2))[1,2]/sum(anova(aov(df1 ~ df2))[,2])

  es_ci <- as.data.frame(ci.pvaf(F.value= F_value, df.1= df_between,
                          df.2= df_resid, N= df_between+df_resid+1, conf.level= .90))

  lower_cl <- es_ci[1,1]
  upper_cl <- es_ci[1,3]

  table <- cbind(F_value, df_between, p_value, eta_squared, lower_cl, upper_cl)

  return(table)
}


# function for chi square table + effect size confidence interval
chix_table <- function(df1, df2){
  # nice chi squared table with cramer's V and CI
  X2 <- chisq.test(df1, df2, correct=FALSE)[1]
  df <- chisq.test(df1, df2, correct=FALSE)[2]
  p_value <- chisq.test(df1, df2, correct=FALSE)[3]

```

```r
  # Craner's V formula in another function (= cv.test)
  cramer_v <- cv.test(df1, df2)

  # transform Cramer's V into Fisher's Z
  fisher_z <- 0.5 * log((1 + cramer_v)/(1 - cramer_v))

  # calculate sample size
  N_size <- sum(table(df1, df2))

  # calculate standard error per Fisher's Z distribution
  standard_error <- 1/sqrt(sum(N_size)-3) * qnorm(1-(0.05/2))

  # confidence interval around Fisher's Z
  ci_fz <- fisher_z + c(-standard_error, standard_error)

  # confidence interval around Cramer's V
  ci_crv <- (exp(2 * ci_fz) - 1)/(1 + exp(2 * ci_fz))

  # bind them all in a table
  table <- cbind(X2, df, p_value, cramer_v, ci_crv[1], ci_crv[2])

  colnames(table)[5] <- "Lower V"
  colnames(table)[6] <- "Upper V"

  table[,c(1,4:6)] <- lapply(table[,c(1,4:6)], function(x){round(x,4)})

  return(table)
}


# function for wilcoxon test table
wilcox_table <-  function(df1, df2){
  # Wilcoxon rank-sum test with ties summary in a nice table

  W <- wilcox.test(df1 ~ df2, correct=TRUE)$statistic
  p_value <- wilcox.test(df1 ~ df2, correct = TRUE)$p.value
  # which method should I use, idk!?!?!?

  cohen_d <- qnorm(p_value)/sqrt(length(df1)) # effect size using Rosenthal's Formula

```

```r
  table <- cbind(W, p_value, cohen_d)

  return(table)

}


# A function to display OR, 95% CL, z score, and p value from the regression summary output (svyolr)
summy <- function(x, round=3) {
  # A function to display OR, CL, and t value from 'svyolr'
  beta <- matrix(coefficients(summary(x))[,1])

  z <- matrix(coefficients(summary(x))[,3])

  p <- matrix(2*pnorm(-abs(z))) # two-tailed z-test

  matrix <- cbind(matrix(beta), matrix(exp(coef(x))), exp(confint(x)), z, p)

  dimnames(matrix)[[2]][1] <- "beta"
  dimnames(matrix)[[2]][2] <- "OR"
  dimnames(matrix)[[2]][5] <- "z-score"
  dimnames(matrix)[[2]][6] <- "p-value"

  return(round(matrix,round))
}


## Function set 4 (custom crosstable)

# Crosstable functions
xt_2x3 <- function(df1, df2){
  # 3x3 table
  main_table <- table(df1, df2)

  r_percent <- round(prop.table(main_table,1),4)

  r1_total <- sum(main_table[1,])
  r2_total <- sum(main_table[2,])

  row_total <- r1_total + r2_total
```

```r

  # row percentage (4th column)
  per_r1 <- round((r1_total/row_total),4)
  per_r2 <- round((r2_total/row_total),4)

  # 4th column
  col4_total <- c(r1_total, per_r1, r2_total, per_r2)

  # column total
  c1_total <- sum(main_table[,1])
  c2_total <- sum(main_table[,2])
  c3_total <- sum(main_table[,3])

  # total N
  total_N <- sum(main_table)

  row5_total <- c(c1_total, c2_total, c3_total, total_N)

  # combine everything into one table
  xtab <- rbind(as.numeric(main_table[1,]), r_percent[1,], as.numeric(main_table[2,]), r_percent
      [2,])

  xtab1 <- cbind(xtab, col4_total)

  xtab2 <- rbind(xtab1, row5_total)

  # rename row and col names
  row_names <- rownames(main_table)

  rownames(xtab2) <- c(row_names[1], "%", row_names[2], "%", "Column Total")
  colnames(xtab2)[4] <- "Row Total"

  return(xtab2)

}

xt_3x3 <- function(df1, df2){
  # 3x3 table
  main_table <- table(df1, df2)

```

```
r_percent <- round(prop.table(main_table,1),4)

r1_total <- sum(main_table[1,])
r2_total <- sum(main_table[2,])
r3_total <- sum(main_table[3,])

row_total <- r1_total + r2_total + r3_total

# row percentage (4th column)
per_r1 <- round((r1_total/row_total),4)
per_r2 <- round((r2_total/row_total),4)
per_r3 <- round((r3_total/row_total),4)

# 4th column
col4_total <- c(r1_total, per_r1, r2_total, per_r2, r3_total, per_r3)

# column total
c1_total <- sum(main_table[,1])
c2_total <- sum(main_table[,2])
c3_total <- sum(main_table[,3])

# total N
total_N <- sum(main_table)

row5_total <- c(c1_total, c2_total, c3_total, total_N)

# combine everything into one table
xtab <- rbind(as.numeric(main_table[1,]), r_percent[1,], as.numeric(main_table[2,]), r_percent
    [2,],
              as.numeric(main_table[3,]), r_percent[3,])

xtab1 <- cbind(xtab, col4_total)

xtab2 <- rbind(xtab1, row5_total)

# rename row and col names
row_names <- rownames(main_table)

rownames(xtab2) <- c(row_names[1],"%", row_names[2], "%", row_names[3], "%", "Column Total")
colnames(xtab2)[4] <- "Row Total"
```

```r

  return(xtab2)

}

xt_4x3 <- function(df1, df2){
  # 4x3 table
  main_table <- table(df1, df2)

  r_percent <- round(prop.table(main_table,1),4)

  r1_total <- sum(main_table[1,])
  r2_total <- sum(main_table[2,])
  r3_total <- sum(main_table[3,])
  r4_total <- sum(main_table[4,])

  row_total <- r1_total + r2_total + r3_total + r4_total

  # row percentage (4th column)
  per_r1 <- round((r1_total/row_total),4)
  per_r2 <- round((r2_total/row_total),4)
  per_r3 <- round((r3_total/row_total),4)
  per_r4 <- round((r4_total/row_total),4)

  # 4th column
  col4_total <- c(r1_total, per_r1, r2_total, per_r2, r3_total, per_r3, r4_total, per_r4)

  # column total
  c1_total <- sum(main_table[,1])
  c2_total <- sum(main_table[,2])
  c3_total <- sum(main_table[,3])

  # total N
  total_N <- sum(main_table)

  row5_total <- c(c1_total, c2_total, c3_total, total_N)

  # combine everything into one table
  xtab <- rbind(as.numeric(main_table[1,]), r_percent[1,], as.numeric(main_table[2,]), r_percent
    [2,],
```

```
                  as.numeric(main_table[3,]), r_percent[3,], as.numeric(main_table[4,]), r_percent
                      [4,])

  xtab1 <- cbind(xtab, col4_total)

  xtab2 <- rbind(xtab1, row5_total)

  # rename row and col names
  row_names <- rownames(main_table)

  rownames(xtab2) <- c(row_names[1],"%", row_names[2], "%", row_names[3], "%", row_names[4], "%", "
      Column Total")
  colnames(xtab2)[4] <- "Row Total"

  return(xtab2)

}

xt_5x3 <- function(df1, df2){
  # 5x3 table
  main_table <- table(df1, df2)

  r_percent <- round(prop.table(main_table,1),4)

  r1_total <- sum(main_table[1,])
  r2_total <- sum(main_table[2,])
  r3_total <- sum(main_table[3,])
  r4_total <- sum(main_table[4,])
  r5_total <- sum(main_table[5,])

  row_total <- r1_total + r2_total + r3_total + r4_total + r5_total

  # row percentage (4th column)
  per_r1 <- round((r1_total/row_total),4)
  per_r2 <- round((r2_total/row_total),4)
  per_r3 <- round((r3_total/row_total),4)
  per_r4 <- round((r4_total/row_total),4)
  per_r5 <- round((r5_total/row_total),4)

  # 4th column
```

```r
  col4_total <- c(r1_total, per_r1, r2_total, per_r2, r3_total, per_r3, r4_total, per_r4, r5_total,
      per_r5)

  # column total
  c1_total <- sum(main_table[,1])
  c2_total <- sum(main_table[,2])
  c3_total <- sum(main_table[,3])

  # total N
  total_N <- sum(main_table)

  row5_total <- c(c1_total, c2_total, c3_total, total_N)

  # combine everything into one table
  xtab <- rbind(as.numeric(main_table[1,]), r_percent[1,], as.numeric(main_table[2,]), r_percent
      [2,],
                as.numeric(main_table[3,]), r_percent[3,], as.numeric(main_table[4,]), r_percent
                    [4,],
                as.numeric(main_table[5,]), r_percent[5,])

  xtab1 <- cbind(xtab, col4_total)

  xtab2 <- rbind(xtab1, row5_total)

  # rename row and col names
  row_names <- rownames(main_table)

  rownames(xtab2) <- c(row_names[1],"%", row_names[2], "%", row_names[3], "%", row_names[4], "%",
      row_names[5], "%", "Column Total")
  colnames(xtab2)[4] <- "Row Total"

  return(xtab2)

}



#### NZHS 2002/03 - 15 years and over ####


```

```r
## remove clothing adjustment, added 1.2 kg
HS02$Weight <- HS02$Weight + 1.2

## recalculate BMI
HS02$bmi <- HS02$Weight/((HS02$Height/100)^2)

## Create a new bmi category using IOTF cut off
# 0 = underweight; 1= average; 2= overweight; 3=obese
HS02 <- HS02 %>%
  mutate(bmic = ifelse(Sex == "M" & Age == 15 & bmi <16.98, 0,
                 ifelse(Sex == "M" & Age == 16 & bmi <17.53, 0,
                 ifelse(Sex == "M" & Age == 17 & bmi <18.04, 0,
                 ifelse(Sex == "F" & Age == 15 & bmi <17.43, 0,
                 ifelse(Sex == "F" & Age == 16 & bmi <17.9, 0,
                 ifelse(Sex == "F" & Age == 17 & bmi <18.24, 0,
                 ifelse(Age >= 18 & bmi <18.5, 0,
                 ifelse(Sex == "M" & Age == 15 & bmi >= 16.98 & bmi <23.28, 1,
                 ifelse(Sex == "M" & Age == 16 & bmi >= 17.53 & bmi <23.89, 1,
                 ifelse(Sex == "M" & Age == 17 & bmi >= 18.04 & bmi <24.46, 1,
                 ifelse(Sex == "F" & Age == 15 & bmi >= 17.43 & bmi <23.89, 1,
                 ifelse(Sex == "F" & Age == 16 & bmi >= 17.90 & bmi <24.34, 1,
                 ifelse(Sex == "F" & Age == 17 & bmi >= 18.24 & bmi <24.70, 1,
                 ifelse(Age >= 18 & bmi >= 18.5 & bmi <25, 1,
                 ifelse(Sex == "M" & Age == 15 & bmi >= 23.28 & bmi <28.32, 2,
                 ifelse(Sex == "M" & Age == 16 & bmi >= 23.89 & bmi <28.89, 2,
                 ifelse(Sex == "M" & Age == 17 & bmi >= 24.46 & bmi <29.43, 2,
                 ifelse(Sex == "F" & Age == 15 & bmi >= 23.89 & bmi <29.01, 2,
                 ifelse(Sex == "F" & Age == 16 & bmi >= 24.34 & bmi <29.40, 2,
                 ifelse(Sex == "F" & Age == 17 & bmi >= 24.70 & bmi <29.70, 2,
                 ifelse(Age >= 18 & bmi >= 25 & bmi <30, 2,
                 ifelse(Sex == "M" & Age == 15 & bmi >=28.32, 3,
                 ifelse(Sex == "M" & Age == 16 & bmi >=28.89, 3,
                 ifelse(Sex == "M" & Age == 17 & bmi >=29.43, 3,
                 ifelse(Sex == "F" & Age == 15 & bmi >=29.01, 3,
                 ifelse(Sex == "F" & Age == 16 & bmi >=29.40, 3,
                 ifelse(Sex == "F" & Age == 17 & bmi >=29.70, 3,
                 ifelse(Age >= 18 & bmi >= 30, 3 ,NA)))))))))))))))))))))))))))))


HS02$bmic <- as.factor(HS02$bmic)
```

```r
summary(HS02$bmic) # 186 underweight


## Exclude underweight
hs02 <- subset(HS02, !bmic == 0 | is.na(bmic))


## Sex
hs02$gender <- factor(hs02$Sex, labels=c("Female","Male"))
hs02$gender <- factor(hs02$gender, levels=c("Male","Female"))

prop_frame(hs02, "gender")


## Age
summary(hs02$Age)

# change the variable name to match the rest of NZHS data
names(hs02)[names(hs02) == "Age"] <- "age"


## Migration Status
# Migrant = less than 10-11 years of residing in NZ
hs02$Q266 <- as.factor(hs02$Q266)

hs02$native <- recode(hs02$Q266, "1993"="Migrant", " 1998" = "Migrant", "1999"="Migrant",
                       " 2000"="Migrant", "2001"="Migrant", "2002"="Migrant",
                      " 2003"="Migrant", "2004"="Migrant", "2005"="Migrant",
                      "2006"="Migrant", "2007"="Migrant", .default="Native")

hs02$native[is.na(hs02$native)] <- "Native"

prop_frame(hs02, "native")



## BMI
# rename to match the rest of NZHS
hs02$bmiscale <- as.numeric(as.character(hs02$bmi))

```

```r
# checking for normality
normal_plot(hs02, "bmiscale")

# label them
hs02$bmic <- factor(hs02$bmic, labels=c("Average","Overweight","Obese"))

prop_frame(hs02, "bmic")


## Ethnicity
# Recode anyone ~  European
hs02 = hs02 %>%
  mutate(euro = ifelse(Q263_01 == 1, 1, 0))

# Recode anyone ~  Maori
hs02 = hs02 %>%
  mutate(maori = ifelse(Q263_02 == 1, 1, 0))

# Recode anyone ~ Pacific
hs02 = hs02 %>%
  mutate(pacific = ifelse(Q263_03 == 1, 1,
                   ifelse(Q263_04 == 1, 1,
                   ifelse(Q263_05 == 1, 1,
                   ifelse(Q263_06 == 1, 1,
                   ifelse(Q263_07 == 1, 1, 0))))))

# Recode anyone ~ asian (chinese and indian only)
hs02 = hs02 %>%
  mutate(asian = ifelse(Q263_08 == 1, 1,
                 ifelse(Q263_09 == 1, 1, 0)))

# Recode anyone ~ other (korena, sri langkan, other asian are here)
hs02 = hs02 %>%
  mutate(other =  ifelse(Q263_10 == 1, 1,
                  ifelse(Q263_11 == 1, 1,
                  ifelse(Q263_12 == 1, 1,
                  ifelse(Q263_13 == 1, 1,
                  ifelse(Q263_14 == 1, 1,
                  ifelse(Q263_Other == 1, 1, 0)))))))
```

```r
# Recode them into: Maori only, Pacific only, Asian only*, European only, Other, 2+ ethnicities (M)
hs02 = hs02 %>%
  mutate(eth_count = ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "Maori Only",
                     ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Pacific Only",
                     ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Asian Only",
                     ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 0, "European Only",
                     ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other",
                     ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+ Ethnicities (M)",
                     ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "2+ Ethnicities (M)",
                     ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Other",
                     ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Other",
                     ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other",
                     ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                     ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                     ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other",
                     ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other",
                     ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other",
                     ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+ Ethnicities (M)",
                     ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                     ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                     ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other",
```

```
                    ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other",
                    ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other",
                    ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                    ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                    ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+ Ethnicities (M)",
                    ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other",
                    ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                    ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                    ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+ Ethnicities (M)",
                    ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other",
                    ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+ Ethnicities (M)",
                    ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+ Ethnicities (M)", NA)))))))))))))))))))))))))))))))

hs02$eth_count <- as.factor(hs02$eth_count)
prop_frame(hs02, "eth_count")


## Household Income
hs02$Q293 <- as.factor(hs02$Q293)
hs02$Q293_imputeflag <- as.factor(hs02$Q293_imputeflag)

# remove imputation
hs02 <- hs02 %>%
  mutate(hhinc = ifelse(Q293_imputeflag == 1, as.numeric(as.character(Q293)), NA))

hs02$hhinc <- as.factor(hs02$hhinc)
summary(hs02$hhinc) # made sure the code works

```

```
# missing data
sum(is.na(hs02$hhinc[hs02$age >=18]))
sum(is.na(hs02$hhinc[hs02$age >=18]))/length(hs02$hhinc[hs02$age >=18])

hs02$hhinc <- as.numeric(as.character(hs02$hhinc))
hs02$hhinc <- cut(hs02$hhinc, breaks=c(0,5,9,11,13))
hs02$hhinc <- factor(hs02$hhinc, labels=c("<=$15,000","$15,001-$40,000","$40,001-$70,000",
                                            ">$70,000"))

prop_frame(hs02, "hhinc")


## Education
# recode 11 = refused to answer into NA
hs02$Q279 <- as.factor(hs02$Q279)
hs02$Q279[hs02$Q279 == "11"] <- NA

# recode secondary qualification
hs02$secondary <- recode(hs02$Q279, "1" = "1", .default = "2") # 1= no degree, 2= secondary school grads


# recode 97 = refused to answer into NA
hs02$Q280 <- as.factor(hs02$Q280)
hs02$Q280[hs02$Q280 == "97"] <- NA # 97: refused to answer

# recode tertiary qualification
hs02$tertiary <- recode(hs02$Q280, "99" = "0", .default = "3") #99 = no tertiary, 3= tertiary
summary(hs02$tertiary)

# recode educational qualification
hs02 <- mutate(hs02, edu = ifelse(secondary =="1" & tertiary =="0", 1,
                         ifelse(secondary == "2" & tertiary == "0", 2,
                         ifelse(tertiary == "3", 3, 1))))

hs02$edu <- as.factor(hs02$edu)
hs02$edu <- factor(hs02$edu, labels=c("No Qualification","Secondary","Tertiary"))

prop_frame(hs02, "edu")

```

```r

## Deprivation Quintile
# recode XX into NA
summary(hs02$Nzdep01_quintiles)
hs02$Nzdep01_quintiles[hs02$Nzdep01_quintiles == "XX"] <- NA

# rearrange
hs02$Nzdep01_quintiles <- factor(hs02$Nzdep01_quintiles, labels=c("1","2","3","4","5"))

prop_frame(hs02, "Nzdep01_quintiles")

# rename
names(hs02)[names(hs02) == "Nzdep01_quintiles"] <- "dep"


## Urban/Rural area
# collapse them into two category
hs02$UA <- recode(hs02$UA_type1, "Rural Centre"="Rural", "Rural"="Rural", .default="Urban")

prop_frame(hs02, "UA")


## Alcohol Problem
hs02$Audit <- as.numeric(as.character(hs02$Audit))

# recode them into alcohol problem status (yes or no)
hs02$haz_drinker_all <- cut(hs02$Audit, breaks= c(-Inf, 7, Inf), labels = c("No Alcohol Problem", "
    Alcohol Problem"))

prop_frame(hs02, "haz_drinker_all")


## Physical Activity
# Q139: how many numbers of days of the last 7 days you were active for
hs02$Q139 <- as.factor(hs02$Q139)
hs02$Q139 <- as.numeric(as.character(hs02$Q139))
hs02$Q139[hs02$Q139 == 8] <- 0    # 8 = no activity

# recode NA into 0 minute/hour
hs02$Q134_hrs[is.na(hs02$Q134_hrs)] <- 0
```

```r
hs02$Q134_mins[is.na(hs02$Q134_mins)] <- 0
hs02$Q136_hrs[is.na(hs02$Q136_hrs)] <- 0
hs02$Q136_mins[is.na(hs02$Q136_mins)] <- 0
hs02$Q138_hrs[is.na(hs02$Q138_hrs)] <- 0
hs02$Q138_mins[is.na(hs02$Q138_mins)] <- 0

# calculate total physical activity duration per week in minutes (rigorous activity received double
    weight)
hs02$minutes <- (60*(hs02$Q134_hrs + hs02$Q136_hrs + 2*(hs02$Q138_hrs)) +
                   (hs02$Q134_mins + hs02$Q136_mins + (2*hs02$Q138_mins)))

# recode them into active or not active
hs02$active <- ifelse(hs02$Q139 >= 5 & hs02$minutes >=30, "Active", "Not Active")
hs02$active <- as.factor(hs02$active)

hs02$active <- factor(hs02$active, levels=c("Not Active", "Active"))

prop_frame(hs02, "active")


## Sedentary Lifestyle
# recode physical activity in minutes into sedentary or not sedentary
hs02$sedentary <- ifelse(hs02$minutes >=30, "Not Sedentary", "Sedentary" )

prop_frame(hs02, "sedentary")


## Difficulty Climbing Several Steps of Stairs
hs02$Q234 <- as.factor(hs02$Q234)

hs02$Q234[hs02$Q234 == 4] <- NA   # 4 = do not know
hs02$Q234[hs02$Q234 == 5] <- NA   # 5 = refused
hs02$stair <- factor(hs02$Q234, labels=c("A Lot Difficult","A Little Difficult", "No Difficulty"))

prop_frame(hs02, "stair")


## Fruit and Vegetable Guideline
## Q156= fruit, Q157= vegetable
hs02$Q156 <- as.factor(hs02$Q156)
```

```r
hs02$Q157 <- as.factor(hs02$Q157)

summary(hs02$Q156)
summary(hs02$Q157)

hs02$Q156[hs02$Q156 == 11] <- NA   # 12 = refused
hs02$Q156[hs02$Q156 == 12] <- NA   # 12 = refused

hs02$Q157[hs02$Q157 == 11] <- NA
hs02$Q157[hs02$Q157 == 12] <- NA

# recode into meeting the serving suggestions of 2+/day for fruit and 3+/day for vegetable
hs02$fruit <- recode(hs02$Q156, "9"="No", "10"="No", "1"="No",
                     .default="Yes")
hs02$fruit <- as.factor(hs02$fruit)

hs02$veges <- recode(hs02$Q157, "3"="Yes", "4"="Yes", .default="No")
hs02$veges <- as.factor(hs02$veges)

prop_frame(hs02, "fruit")
prop_frame(hs02, "veges")


## Rename cluster and strata so that they will merge with the rest of the NZHS
names(hs02)[names(hs02) == 'PSU_no'] <- 'cluster'
names(hs02)[names(hs02) == 'Stratum_no'] <- 'strata'
names(hs02)[names(hs02) == 'Finalwgt'] <- 'finalwgt'



#### NZHS 2006/07 - adult ####

## Transform BMI into numeric, removing character strings
summary(HS06A$BMI) # 407 Unknown (U) and 401 Refused (R)
colnames(HS06A)[which(names(HS06A) == "BMI")] <- "bmi"
HS06A$bmiscale <- as.numeric(as.character(HS06A$bmi))
summary(HS06A$bmiscale)


## Recode 95+ in age variable into 95 (the max value is censored at this number)
```

```
HS06A$age <- as.character(HS06A$age)
HS06A$age[HS06A$age == "95+"] <- 95
HS06A$age <- as.numeric(HS06A$age)

HS06A$D_01 <- as.factor(HS06A$D_01) # 1=male, 8=female

## Recode BMI category using IOTF cut-offs
# 0 = underweight; 1= average; 2= overweight; 3=obese
HS06A <- HS06A %>%
  mutate(bmic = ifelse(D_01 == 1 & age == 15 & bmiscale <16.98, 0,
                ifelse(D_01 == 1 & age == 16 & bmiscale <17.53, 0,
                ifelse(D_01 == 1 & age == 17 & bmiscale <18.04, 0,
                ifelse(D_01 == 8 & age == 15 & bmiscale <17.43, 0,
                ifelse(D_01 == 8 & age == 16 & bmiscale <17.9, 0,
                ifelse(D_01 == 8 & age == 17 & bmiscale <18.24, 0,
                ifelse(age >= 18 & bmiscale <18.5, 0,
                ifelse(D_01 == 1 & age == 15 & bmiscale >= 16.98 & bmiscale <23.28, 1,
                ifelse(D_01 == 1 & age == 16 & bmiscale >= 17.53 & bmiscale <23.89, 1,
                ifelse(D_01 == 1 & age == 17 & bmiscale >= 18.04 & bmiscale <24.46, 1,
                ifelse(D_01 == 8 & age == 15 & bmiscale >= 17.43 & bmiscale <23.89, 1,
                ifelse(D_01 == 8 & age == 16 & bmiscale >= 17.90 & bmiscale <24.34, 1,
                ifelse(D_01 == 8 & age == 17 & bmiscale >= 18.24 & bmiscale <24.70, 1,
                ifelse(age >= 18 & bmiscale >= 18.5 & bmiscale <25, 1,
                ifelse(D_01 == 1 & age == 15 & bmiscale >= 23.28 & bmiscale <28.32, 2,
                ifelse(D_01 == 1 & age == 16 & bmiscale >= 23.89 & bmiscale <28.89, 2,
                ifelse(D_01 == 1 & age == 17 & bmiscale >= 24.46 & bmiscale <29.43, 2,
                ifelse(D_01 == 8 & age == 15 & bmiscale >= 23.89 & bmiscale <29.01, 2,
                ifelse(D_01 == 8 & age == 16 & bmiscale >= 24.34 & bmiscale <29.40, 2,
                ifelse(D_01 == 8 & age == 17 & bmiscale >= 24.70 & bmiscale <29.70, 2,
                ifelse(age >= 18 & bmiscale >= 25 & bmiscale <30, 2,
                ifelse(D_01 == 1 & age == 15 & bmiscale >=28.32, 3,
                ifelse(D_01 == 1 & age == 16 & bmiscale >=28.89, 3,
                ifelse(D_01 == 1 & age == 17 & bmiscale >=29.43, 3,
                ifelse(D_01 == 8 & age == 15 & bmiscale >=29.01, 3,
                ifelse(D_01 == 8 & age == 16 & bmiscale >=29.40, 3,
                ifelse(D_01 == 8 & age == 17 & bmiscale >=29.70, 3,
                ifelse(age >= 18 & bmiscale >= 30, 3 ,NA)))))))))))))))))))))))))))))

HS06A$bmic <- as.factor(HS06A$bmic)
summary(HS06A$bmic) # 139 underweight, 808 NA
```

```r


# Subset the data, remove underweight
a06 <- subset(HS06A, bmic != 0 | is.na(bmic))

# label
a06$bmic <- factor(a06$bmic, labels=c("Average", "Overweight","Obese"))
summary(a06$bmic)


## Sex (D_01) (1 = male, 8 = female)
sum(is.na(a06$D_01))
a06$gender <- factor(a06$D_01, labels = c("Male", "Female"))

prop_frame(a06, "gender")


## Age
summary(a06$age)


## Migration Status
summary(a06$a5_05)

a06$native <- recode(a06$a5_05, "1997"="Migrant", " 1998" = "Migrant", "1999"="Migrant",
                     " 2000"="Migrant", "2001"="Migrant", "2002"="Migrant", " 2003"="Migrant",
                     "2004"="Migrant", "2005"="Migrant", "2006"="Migrant", "2007"="Migrant",
                     "K"= "No", "R"= "No", .default="Native") # K and R response are treated as NA

a06$native[is.na(a06$native)] <- "Native"
a06$native[a06$native == "No"] <- NA
a06$native <- as.factor(a06$native)

prop_frame(a06, "native")


## BMI
summary(a06$bmiscale)

prop_frame(a06, "bmic")
```

```
## Ethnicity
a06$A5_02_group_1 <- as.factor(a06$A5_02_group_1)
a06$A5_02_group_2 <- as.factor(a06$A5_02_group_2)
a06$A5_02_group_3 <- as.factor(a06$A5_02_group_3)
a06$A5_02_group_4 <- as.factor(a06$A5_02_group_4)
a06$A5_02_group_5 <- as.factor(a06$A5_02_group_5)


# Recode anyone ~ European
a06 <- a06 %>%
  mutate(euro = ifelse(A5_02_group_1 %in% c(1,9), 1,
                ifelse(A5_02_group_2 %in% c(1,9), 1,
                ifelse(A5_02_group_3 %in% c(1,9), 1, 0))))

# Recode anyone ~ Maori
a06 <- a06 %>%
  mutate(maori = ifelse(A5_02_group_1 %in% 2, 1,
                ifelse(A5_02_group_2 %in% 2, 1,
                ifelse(A5_02_group_3 %in% 2, 1,
                ifelse(A5_02_group_4 %in% 2, 1, 0)))))

# Recode anyone ~ Pacific
a06 <- a06 %>%
  mutate(pacific = ifelse(A5_02_group_1 %in% c(3,4,5,6,10), 1,
                ifelse(A5_02_group_2 %in% c(3,4,5,6,10), 1,
                ifelse(A5_02_group_3 %in% c(3,4,5,6,10), 1,
                ifelse(A5_02_group_4 %in% c(3,4,5,6,10), 1,
                ifelse(A5_02_group_5 %in% c(3,4,5,6,10), 1, 0))))))

# Recode anyone ~ Asian (Indian and Chinese only)
a06 <- a06 %>%
  mutate(asian = ifelse(A5_02_group_1 %in% c(7,8), 1,
                ifelse(A5_02_group_2 %in% c(7,8), 1,
                ifelse(A5_02_group_3 %in% c(7,8), 1,
                ifelse(A5_02_group_4 %in% c(7,8), 1,
                ifelse(A5_02_group_5 %in% c(7,8), 1, 0))))))

# Recode anyone ~ Other (other asian is here)
```

```r
a06 <- a06 %>%
  mutate(other = ifelse(A5_02_group_1 %in% c(11,12), 1,
                ifelse(A5_02_group_2 %in% c(11,12), 1,
                ifelse(A5_02_group_3 %in% c(11,12), 1,
                ifelse(A5_02_group_4 %in% c(11,12), 1,
                ifelse(A5_02_group_5 %in% c(11,12), 1, 0))))))


# Recode them into: Maori only, Pacific only, Asian only*, European only, Other, 2+ ethnicities (M)
a06 = a06 %>%
  mutate(eth_count = ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "Maori Only",
                     ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Pacific Only",
                     ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Asian Only",
                     ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 0, "European Only",
                     ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other",
                     ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+ Ethnicities (M)",
                     ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "2+ Ethnicities (M)",
                     ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Other",
                     ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Other",
                     ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other",
                     ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                     ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                     ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other",
                     ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other",
                     ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other",
```

```
                    ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+ Ethnicities (M)",
                    ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                    ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                    ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other",
                    ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other",
                    ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other",
                    ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                    ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                    ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+ Ethnicities (M)",
                    ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other",
                    ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                    ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                    ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+ Ethnicities (M)",
                    ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other",
                    ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+ Ethnicities (M)",
                    ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+ Ethnicities (M)", NA)))))))))))))))))))))))))))))))

a06$eth_count <- as.factor(a06$eth_count)

prop_frame(a06, "eth_count")


## Deprivation Quintile
a06$nzdep06_quintile <- as.factor(a06$nzdep06_quintile)
```

```r

names(a06)[names(a06) == "nzdep06_quintile"] <- 'dep'

prop_frame(a06, "dep")


## Fruit and Vegetable
# 28 = fruit, 29 = vegetable
a06$A3_28 <- as.factor(a06$A3_28)
a06$A3_29 <- as.factor(a06$A3_29)
summary(a06$A3_28)
summary(a06$A3_29) # 1 answered "don't know"

a06$A3_29[a06$A3_29 =="K"] <- NA

# Adherence to fruit and vegetable guideline
a06$fruit <- as.numeric(as.character(a06$A3_28))
a06$fruit <- cut(a06$fruit, breaks=c(0,3,6))
summary(a06$fruit)

a06$fruit <- factor(a06$fruit, labels=c("No","Yes"))

a06$veges <- as.numeric(as.character(a06$A3_28))
a06$veges <- cut(a06$veges, breaks=c(0,4,6))
summary(a06$veges)

a06$veges <- factor(a06$veges, labels=c("No","Yes"))

prop_frame(a06, "fruit")
prop_frame(a06, "veges")


## Smoking Status
a06 <- a06 %>%
  mutate(smoke = ifelse(ex_smoker == 0 & current_smoker == 0, "Non Smoker",
                 ifelse(ex_smoker == 1 & current_smoker == 0, "Ex Smoker", "Current Smoker")))

a06$smoke <- as.factor(a06$smoke)
a06$smoke <- factor(a06$smoke, levels = c("Non Smoker","Ex Smoker", "Current Smoker"))

```

```r
prop_frame(a06, "smoke")


## Alcohol Problem
a06$AUDIT_cat[a06$AUDIT_cat == "K"] <- NA
a06$AUDIT_cat[a06$AUDIT_cat == "R"] <- NA

a06$haz_drinker_all <- recode(a06$AUDIT_cat, "0"="No Alcohol Problem", "1"="Alcohol Problem", .default= NULL)

prop_frame(a06, "haz_drinker_all")


## Urban/Rural Area
a06$UA <- as.factor(a06$UR_06)
a06$UA <- recode(a06$UA, "4"="Rural", .default ="Urban")
a06$UA <- as.factor(a06$UA)

prop_frame(a06, "UA")


## Difficulty Climbing Several Steps of Stair
a06$stair <- as.factor(a06$A4_22)
a06$stair[a06$stair == "K"] <- NA

a06$stair <- factor(a06$stair, labels=c("A Lot Difficult","A Little Difficult","No Difficulty"))

prop_frame(a06, "stair")


## Physical Activity
a06$A3_12 <- as.numeric(as.character(a06$A3_12))
a06$A3_14 <- as.numeric(as.character(a06$A3_14))
a06$A3_16 <- as.numeric(as.character(a06$A3_16))
a06$A3_13a <- as.numeric(as.character(a06$A3_13a))
a06$A3_15a <- as.numeric(as.character(a06$A3_15a))
a06$A3_17a <- as.numeric(as.character(a06$A3_17a))
a06$A3_13b <- as.numeric(as.character(a06$A3_13b))
a06$A3_15b <- as.numeric(as.character(a06$A3_15b))
a06$A3_17b <- as.numeric(as.character(a06$A3_17b))
```

```r
a06 <- mutate(a06, minutes= (60*(A3_13a+A3_15a+2*(A3_17a)))+A3_13b+A3_15b+(2*A3_17b))
summary(a06$minutes) # physical activity in minutes in the past week

# active at least 30' per day for 5 days in the last 1 week (set question)
a06$active <- as.numeric(as.character(a06$A3_18))
a06$active <- cut(a06$active, breaks=c(-Inf,4,Inf), labels=c("Not Active","Active"))

prop_frame(a06, "active")

## Sedentary Lifestyle
a06 <- mutate(a06, sedentary = ifelse(minutes <30, "Sedentary", "Not Sedentary"))

a06$sedentary <- as.factor(a06$sedentary)

prop_frame(a06, "sedentary")


## Household Income
a06$hhinc <- ordered(a06$A5_24, levels=c("1","2","3","4","5","6","7","8","9","10","11","12","13","14","15"))
sum(is.na(a06$hhinc)) # 1450 missing
sum(is.na(a06$hhinc))/length(a06$hhinc) # 12.5639% missing

a06$hhinc <- as.numeric(as.character(a06$hhinc))
a06$hhinc <- cut(a06$hhinc, breaks=c(0,3,7,10,15))
a06$hhinc <- factor(a06$hhinc, labels=c("<=$15,000","$15,001-$40,000","$40,001-$70,000",">$70,000"))

prop_frame(a06, "hhinc")


## Education
a06$secondary <- recode(a06$A5_13, "1" = "1", .default = "2") # 1= no degree, 2= secondary school grads

a06$tertiary <- recode(a06$A5_14_group, "1" = "0", .default = "3") # 0 = no tertiary, 3= tertiary

a06 <- mutate(a06, edu = ifelse(secondary =="1" & tertiary =="0", 1,
                         ifelse(secondary == "2" & tertiary == "0", 2,
                         ifelse(tertiary == "3", 3, 1))))

```

```
a06$edu <- as.factor(a06$edu)
a06$edu <- factor(a06$edu, labels=c("No Qualification","Secondary","Tertiary"))

prop_frame(a06, "edu")


## Recode unique clusters for each stratum
names(a06)[names(a06) == "stratum"] <- "strata"

summary(a06$cluster) # min 1, max 137; start recoding with 138

# create a code book for a new cluster coding
l <- c(1:137) # Var1 = cluster
m <- c(1:21) # Var2 = strata

k <- data.frame(expand.grid(l, m))
k$new.cluster <- c(1:nrow(k))

# create a backup of original clusters
a06$clu <- a06$cluster

# recode them into new unique clusters
for (i in 1:137){
  for (j in 1:21){
    a06$cluster[a06$cluster == i & a06$strata == j] <- k$new.cluster[k$Var1 == i & k$Var2 == j]
  }
}

# save object "k" to use it in the child data set



#### NZHS 2006/07 - child ####

summary(HS06C$BMI) # 148 refused, 8 unknown
colnames(HS06C)[which(names(HS06C) == "BMI")] <- "bmi"

# remove non-numeric data
HS06C$bmiscale <- (as.numeric(as.character(HS06C$bmi)))

```

```
HS06C <- HS06C %>%
  mutate(under =
  ifelse(sex == "M" & age == 2 & bmiscale <15.24, 0,
  ifelse(sex == "M" & age == 3 & bmiscale <14.83, 0,
  ifelse(sex == "M" & age == 4 & bmiscale <14.51, 0,
  ifelse(sex == "M" & age == 5 & bmiscale <14.26, 0,
  ifelse(sex == "M" & age == 6 & bmiscale <14.06, 0,
  ifelse(sex == "M" & age == 7 & bmiscale <14.00, 0,
  ifelse(sex == "M" & age == 8 & bmiscale <14.13, 0,
  ifelse(sex == "M" & age == 9 & bmiscale <14.36, 0,
  ifelse(sex == "M" & age == 10 & bmiscale <14.63, 0,
  ifelse(sex == "M" & age == 11 & bmiscale <14.96, 0,
  ifelse(sex == "M" & age == 12 & bmiscale <15.36, 0,
  ifelse(sex == "M" & age == 13 & bmiscale <15.84, 0,
  ifelse(sex == "M" & age == 14 & bmiscale <16.39, 0,
  ifelse(sex == "M" & age == 15 & bmiscale <16.98, 0,
  ifelse(sex == "M" & age == 16 & bmiscale <17.53, 0,
  ifelse(sex == "M" & age == 17 & bmiscale <18.04, 0,
  ifelse(sex == "F" & age == 2 & bmiscale <14.96, 0,
  ifelse(sex == "F" & age == 3 & bmiscale <14.60, 0,
  ifelse(sex == "F" & age == 4 & bmiscale <14.30, 0,
  ifelse(sex == "F" & age == 5 & bmiscale <14.04, 0,
  ifelse(sex == "F" & age == 6 & bmiscale <13.85, 0,
  ifelse(sex == "F" & age == 7 & bmiscale <13.83, 0,
  ifelse(sex == "F" & age == 8 & bmiscale <14.00, 0,
  ifelse(sex == "F" & age == 9 & bmiscale <14.26, 0,
  ifelse(sex == "F" & age == 10 & bmiscale <14.58, 0,
  ifelse(sex == "F" & age == 11 & bmiscale <15.03, 0,
  ifelse(sex == "F" & age == 12 & bmiscale <15.59, 0,
  ifelse(sex == "F" & age == 13 & bmiscale <16.23, 0,
  ifelse(sex == "F" & age == 14 & bmiscale <16.86, 0,
  ifelse(sex == "F" & age == 15 & bmiscale <17.43, 0,
  ifelse(sex == "F" & age == 16 & bmiscale <17.90, 0,
  ifelse(sex == "F" & age == 17 & bmiscale <18.24, 0,
  ifelse(age >= 18 & bmiscale <18.5, 0, NA))))))))))))))))))))))))))))))))))

HS06C <- HS06C %>%
  mutate(average =
  ifelse(sex == "M" & age == 2 & bmiscale >=15.24 & bmiscale <18.36, 1,
  ifelse(sex == "M" & age == 3 & bmiscale >=14.83 & bmiscale <17.85, 1,
```

```
  ifelse(sex == "M" & age == 4 & bmiscale >=14.51 & bmiscale <17.52, 1,
  ifelse(sex == "M" & age == 5 & bmiscale >=14.26 & bmiscale <17.39, 1,
  ifelse(sex == "M" & age == 6 & bmiscale >=14.06 & bmiscale <17.52, 1,
  ifelse(sex == "M" & age == 7 & bmiscale >=14.00 & bmiscale <17.88, 1,
  ifelse(sex == "M" & age == 8 & bmiscale >=14.13 & bmiscale <18.41, 1,
  ifelse(sex == "M" & age == 9 & bmiscale >=14.36 & bmiscale <19.07, 1,
  ifelse(sex == "M" & age == 10 & bmiscale >=14.63 & bmiscale <19.80, 1,
  ifelse(sex == "M" & age == 11 & bmiscale >=14.96 & bmiscale <20.51, 1,
  ifelse(sex == "M" & age == 12 & bmiscale >=15.36 & bmiscale <21.20, 1,
  ifelse(sex == "M" & age == 13 & bmiscale >=15.84 & bmiscale <21.89, 1,
  ifelse(sex == "M" & age == 14 & bmiscale >=16.39 & bmiscale <22.60, 1,
  ifelse(sex == "M" & age == 15 & bmiscale >=16.98 & bmiscale <23.28, 1,
  ifelse(sex == "M" & age == 16 & bmiscale >=17.53 & bmiscale <23.89, 1,
  ifelse(sex == "M" & age == 17 & bmiscale >=18.04 & bmiscale <24.46, 1,
  ifelse(sex == "F" & age == 2 & bmiscale >=14.96 & bmiscale <18.09, 1,
  ifelse(sex == "F" & age == 3 & bmiscale >=14.60 & bmiscale <17.64, 1,
  ifelse(sex == "F" & age == 4 & bmiscale >=14.30 & bmiscale <17.35, 1,
  ifelse(sex == "F" & age == 5 & bmiscale >=14.04 & bmiscale <17.23, 1,
  ifelse(sex == "F" & age == 6 & bmiscale >=13.85 & bmiscale <17.33, 1,
  ifelse(sex == "F" & age == 7 & bmiscale >=13.83 & bmiscale <17.69, 1,
  ifelse(sex == "F" & age == 8 & bmiscale >=14.00 & bmiscale <18.28, 1,
  ifelse(sex == "F" & age == 9 & bmiscale >=14.26 & bmiscale <18.99, 1,
  ifelse(sex == "F" & age == 10 & bmiscale >=14.58 & bmiscale <19.78, 1,
  ifelse(sex == "F" & age == 11 & bmiscale >=15.03 & bmiscale <20.66, 1,
  ifelse(sex == "F" & age == 12 & bmiscale >=15.59 & bmiscale <21.59, 1,
  ifelse(sex == "F" & age == 13 & bmiscale >=16.23 & bmiscale <22.49, 1,
  ifelse(sex == "F" & age == 14 & bmiscale >=16.86 & bmiscale <23.27, 1,
  ifelse(sex == "F" & age == 15 & bmiscale >=17.43 & bmiscale <23.89, 1,
  ifelse(sex == "F" & age == 16 & bmiscale >=17.90 & bmiscale <24.34, 1,
  ifelse(sex == "F" & age == 17 & bmiscale >=18.24 & bmiscale <24.70, 1,
  ifelse(age >= 18 & bmiscale >=18.5 & bmiscale <25, 1, NA))))))))))))))))))))))))))))))))))

HS06C <- HS06C %>%
  mutate(over=
  ifelse(sex == "M" & age == 2 & bmiscale >=18.36 & bmiscale <19.99, 2,
  ifelse(sex == "M" & age == 3 & bmiscale >=17.85 & bmiscale <19.50, 2,
  ifelse(sex == "M" & age == 4 & bmiscale >=17.52 & bmiscale <19.23, 2,
  ifelse(sex == "M" & age == 5 & bmiscale >=17.39 & bmiscale <19.27, 2,
  ifelse(sex == "M" & age == 6 & bmiscale >=17.52 & bmiscale <19.76, 2,
  ifelse(sex == "M" & age == 7 & bmiscale >=17.88 & bmiscale <20.59, 2,
```

```
  ifelse(sex == "M" & age == 8 & bmiscale >=18.41 & bmiscale <21.56, 2,
  ifelse(sex == "M" & age == 9 & bmiscale >=19.07 & bmiscale <22.71, 2,
  ifelse(sex == "M" & age == 10 & bmiscale >=19.80 & bmiscale <23.96, 2,
  ifelse(sex == "M" & age == 11 & bmiscale >=20.51 & bmiscale <25.07, 2,
  ifelse(sex == "M" & age == 12 & bmiscale >=21.20 & bmiscale <26.02, 2,
  ifelse(sex == "M" & age == 13 & bmiscale >=21.89 & bmiscale <26.87, 2,
  ifelse(sex == "M" & age == 14 & bmiscale >=22.60 & bmiscale <27.64, 2,
  ifelse(sex == "M" & age == 15 & bmiscale >=23.28 & bmiscale <28.32, 2,
  ifelse(sex == "M" & age == 16 & bmiscale >=23.89 & bmiscale <28.89, 2,
  ifelse(sex == "M" & age == 17 & bmiscale >=24.46 & bmiscale <29.43, 2,
  ifelse(sex == "F" & age == 2 & bmiscale >=18.09 & bmiscale <19.81, 2,
  ifelse(sex == "F" & age == 3 & bmiscale >=17.64 & bmiscale <19.38, 2,
  ifelse(sex == "F" & age == 4 & bmiscale >=17.35 & bmiscale <19.16, 2,
  ifelse(sex == "F" & age == 5 & bmiscale >=17.23 & bmiscale <19.20, 2,
  ifelse(sex == "F" & age == 6 & bmiscale >=17.33 & bmiscale <19.61, 2,
  ifelse(sex == "F" & age == 7 & bmiscale >=17.69 & bmiscale <20.39, 2,
  ifelse(sex == "F" & age == 8 & bmiscale >=18.28 & bmiscale <21.44, 2,
  ifelse(sex == "F" & age == 9 & bmiscale >=18.99 & bmiscale <22.66, 2,
  ifelse(sex == "F" & age == 10 & bmiscale >=19.78 & bmiscale <23.97, 2,
  ifelse(sex == "F" & age == 11 & bmiscale >=20.66 & bmiscale <25.25, 2,
  ifelse(sex == "F" & age == 12 & bmiscale >=21.59 & bmiscale <26.47, 2,
  ifelse(sex == "F" & age == 13 & bmiscale >=22.49 & bmiscale <27.57, 2,
  ifelse(sex == "F" & age == 14 & bmiscale >=23.27 & bmiscale <28.42, 2,
  ifelse(sex == "F" & age == 15 & bmiscale >=23.89 & bmiscale <29.01, 2,
  ifelse(sex == "F" & age == 16 & bmiscale >=24.34 & bmiscale <29.40, 2,
  ifelse(sex == "F" & age == 17 & bmiscale >=24.70 & bmiscale <29.70, 2,
  ifelse(age >= 18 & bmiscale >=25 & bmiscale <30, 2, NA))))))))))))))))))))))))))))))))

HS06C <- HS06C %>%
  mutate(obes=
  ifelse(sex == "M" & age == 2 & bmiscale >=19.99, 3,
  ifelse(sex == "M" & age == 3 & bmiscale >=19.50, 3,
  ifelse(sex == "M" & age == 4 & bmiscale >=19.23, 3,
  ifelse(sex == "M" & age == 5 & bmiscale >=19.27, 3,
  ifelse(sex == "M" & age == 6 & bmiscale >=19.76, 3,
  ifelse(sex == "M" & age == 7 & bmiscale >=20.59, 3,
  ifelse(sex == "M" & age == 8 & bmiscale >=21.56, 3,
  ifelse(sex == "M" & age == 9 & bmiscale >=22.71, 3,
  ifelse(sex == "M" & age == 10 & bmiscale >=23.96, 3,
  ifelse(sex == "M" & age == 11 & bmiscale >=25.07, 3,
```

```r
  ifelse(sex == "M" & age == 12 & bmiscale >=26.02, 3,
  ifelse(sex == "M" & age == 13 & bmiscale >=26.87, 3,
  ifelse(sex == "M" & age == 14 & bmiscale >=27.64, 3,
  ifelse(sex == "M" & age == 15 & bmiscale >=28.32, 3,
  ifelse(sex == "M" & age == 16 & bmiscale >=28.89, 3,
  ifelse(sex == "M" & age == 17 & bmiscale >=29.43, 3,
  ifelse(sex == "F" & age == 2 & bmiscale >=19.81, 3,
  ifelse(sex == "F" & age == 3 & bmiscale >=19.38, 3,
  ifelse(sex == "F" & age == 4 & bmiscale >=19.16, 3,
  ifelse(sex == "F" & age == 5 & bmiscale >=19.20, 3,
  ifelse(sex == "F" & age == 6 & bmiscale >=19.61, 3,
  ifelse(sex == "F" & age == 7 & bmiscale >=20.39, 3,
  ifelse(sex == "F" & age == 8 & bmiscale >=21.44, 3,
  ifelse(sex == "F" & age == 9 & bmiscale >=22.66, 3,
  ifelse(sex == "F" & age == 10 & bmiscale >=23.97, 3,
  ifelse(sex == "F" & age == 11 & bmiscale >=25.25, 3,
  ifelse(sex == "F" & age == 12 & bmiscale >=26.47, 3,
  ifelse(sex == "F" & age == 13 & bmiscale >=27.57, 3,
  ifelse(sex == "F" & age == 14 & bmiscale >=28.42, 3,
  ifelse(sex == "F" & age == 15 & bmiscale >=29.01, 3,
  ifelse(sex == "F" & age == 16 & bmiscale >=29.40, 3,
  ifelse(sex == "F" & age == 17 & bmiscale >=29.70, 3,
  ifelse(age >= 18 & bmiscale >=30, 3, NA))))))))))))))))))))))))))))))))


HS06C <- HS06C %>%
  mutate(bmic = ifelse(under == 0 & is.na(average) & is.na(over) & is.na(obes), 0,
                ifelse(average == 1 & is.na(under) & is.na(over) & is.na(obes), 1,
                ifelse(over == 2 & is.na(under) & is.na(average) & is.na(obes), 2,
                ifelse(obes == 3 & is.na(under) & is.na(over) & is.na(average), 3, NA)))))

HS06C$bmic <- as.factor(HS06C$bmic)
summary(HS06C$bmic) # 134 underweight, 903 NA


# exclude below 2 years old and underweight
c06 <- subset(HS06C, age>=2 & bmic != 0 | is.na(bmic))


## BMI
```

```
c06$bmic <- factor(c06$bmic, labels=c("Average","Overweight","Obese"))

prop_frame(c06, "bmic")


## Sex
c06$gender <- factor(c06$sex, labels = c("Female","Male"))
c06$gender <- factor(c06$gender, levels = c("Male","Female"))

prop_frame(c06, "gender")


## Age
summary(c06$age)


## Ethnicity
# recode anyone ~ european
c06 <- c06 %>%
  mutate(euro = ifelse(C4_02_group_1 %in% c(1,9), 1,
                ifelse(C4_02_group_2 %in% c(1,9), 1,
                ifelse(C4_02_group_3 %in% c(1,9), 1, 0))))

# recode anyone ~ maori
c06 <- c06 %>%
  mutate(maori = ifelse(C4_02_group_1 %in% 2, 1,
                ifelse(C4_02_group_2 %in% 2, 1,
                ifelse(C4_02_group_3 %in% 2, 1,
                ifelse(C4_02_group_4 %in% 2, 1, 0)))))

# recode anyone ~ Pacific
c06 <- c06 %>%
  mutate(pacific = ifelse(C4_02_group_1 %in% c(3,4,5,6,10), 1,
                ifelse(C4_02_group_2 %in% c(3,4,5,6,10), 1,
                ifelse(C4_02_group_3 %in% c(3,4,5,6,10), 1,
                ifelse(C4_02_group_4 %in% c(3,4,5,6,10), 1,
                ifelse(C4_02_group_5 %in% c(3,4,5,6,10), 1,
                ifelse(C4_02_group_6 %in% c(3,4,5,6,10), 1,
                ifelse(C4_02_group_7 %in% c(3,4,5,6,10), 1, 0))))))))

```

```r
# recode anyone ~ Asian (Indian and Chinese only)
c06 <- c06 %>%
  mutate(asian = ifelse(C4_02_group_1 %in% c(7,8), 1,
                 ifelse(C4_02_group_2 %in% c(7,8), 1,
                 ifelse(C4_02_group_3 %in% c(7,8), 1,
                 ifelse(C4_02_group_4 %in% c(7,8), 1,
                 ifelse(C4_02_group_5 %in% c(7,8), 1,
                 ifelse(C4_02_group_6 %in% c(7,8), 1,
                 ifelse(C4_02_group_7 %in% c(7,8), 1, 0))))))))

# recode anyone ~ Other (other asian is here)
c06 <- c06 %>%
  mutate(other = ifelse(C4_02_group_1 %in% c(11,12), 1,
                 ifelse(C4_02_group_2 %in% c(11,12), 1,
                 ifelse(C4_02_group_3 %in% c(11,12), 1,
                 ifelse(C4_02_group_4 %in% c(11,12), 1,
                 ifelse(C4_02_group_5 %in% c(11,12), 1,
                 ifelse(C4_02_group_6 %in% c(11,12), 1,
                 ifelse(C4_02_group_7 %in% c(11,12), 1, 0))))))))


# Recode them into: Maori only, Pacific only, Asian only, European only, Other, 2+ ethnicities (M)
c06 = c06 %>%
  mutate(eth_count = ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "Maori Only",
                     ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Pacific Only",
                     ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Asian Only",
                     ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 0, "European Only",
                     ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other",
                     ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+ Ethnicities (M)",
                     ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "2+ Ethnicities (M)",
                     ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Other",
```

```
ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Other",
ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other",
ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+ Ethnicities (M)",
ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+ Ethnicities (M)",
ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other",
ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other",
ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other",
ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+ Ethnicities (M)",
ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+ Ethnicities (M)",
ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+ Ethnicities (M)",
ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other",
ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other",
ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other",
ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+ Ethnicities (M)",
ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+ Ethnicities (M)",
ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+ Ethnicities (M)",
ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other",
ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+ Ethnicities (M)",
ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+ Ethnicities (M)",
ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+ Ethnicities (M)",
```

```
                        ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other",
                        ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+ Ethnicities (M)",
                        ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+ Ethnicities (M)", NA)))))))))))))))))))))))))))))))

c06$eth_count <- as.factor(c06$eth_count)

prop_frame(c06, "eth_count")


## Urban/Rural Area
c06$UA <- recode(c06$ur_desc, "Rural"="Rural", .default="Urban")

prop_frame(c06, "UA")


## Household Income
summary(c06$C4_22)
c06$hhinc <- ordered(c06$C4_22, levels=c("1","2","3","4","5","6","7","8","9","10","11","12",
                                          "13","14","15"))
sum(is.na(c06$hhinc))
sum(is.na(c06$hhinc))/length(c06$hhinc) # 369 NA, 9.5% missing

c06$hhinc <- as.numeric(as.character(c06$hhinc))
c06$hhinc <- cut(c06$hhinc, breaks=c(0,3,7,10,15))
c06$hhinc <- factor(c06$hhinc, labels=c("<=$15,000","$15,001-$40,000","$40,001-$70,000",">$70,000"))

prop_frame(c06, "hhinc")


## Education
c06$edu1 <- as.numeric(as.character(c06$C4_19))
c06$secondary <- recode(c06$edu1, "1" = "1", .default = "2") # 1= no degree, 2= secondary school grads

summary(c06$C4_20_group)
c06$edu2 <- as.numeric(as.character(c06$C4_20_group))
c06$tertiary <- recode(c06$edu2, "1" = "0", .default = "3") #0 = no tertiary, 3= tertiary
```

```r
# prioritise highest educational qualification in the household
c06 <- mutate(c06, edu = ifelse(secondary =="1" & tertiary =="0", 1,
                        ifelse(secondary == "2" & tertiary == "0", 2,
                        ifelse(tertiary == "3", 3, 1))))

c06$edu <- as.factor(c06$edu)
c06$edu <- factor(c06$edu, labels=c("No Qualification","Secondary","Tertiary"))

prop_frame(c06, "edu")


## Deprivation Quintile
names(c06)[names(c06) == "nzdep06_quintile"] <- "dep"

prop_frame(c06, "dep")


## Soft Drink and Fast Food
# categorise the soft drink consumption to (0 = 0/week, 1 = 1/week, 2 = 2-3/week, 3 = 4+/week)
c06$C3_13 <- as.factor(c06$C3_13)
c06$softd <- c06$C3_13
c06$softd <- recode(c06$softd, "0"="0/week","1"="1/week" , "2" = "2-3/week", "3" = "2-3/week",
                    .default="4+/week")
c06$softd <- ordered(c06$softd, levels = c("0/week", "1/week", "2-3/week", "4+/week"))

prop_frame(c06, "softd")


# categorise fast food consumption to (0 = 0/week, 1 = 1/week, 2 = 2-3/week, 3 = 4+/week)
c06$fastf <- c06$C3_15
c06$fastf <- recode(c06$fastf, "0"="0/week", "1"="1/week", "2"="2-3/week",
                    "3"="2-3/week", .default="4+/week")
c06$fastf <- ordered(c06$fastf, levels =c("0/week", "1/week", "2-3/week", "4+/week"))

prop_frame(c06, "fastf")


## Recode Unique Clusters
names(c06)[names(c06) == "stratum"] <- "strata"
```

```
# create backup
c06$clu <- c06$cluster

# recode to unique clusters using the same codebook (k) as the one in adult data set
for (i in 1:131){
  for (j in 1:21){
    c06$cluster[c06$cluster == i & c06$strata == j] <- k$new.cluster[k$Var1 == i & k$Var2 == j]
  }
}

summary(c06$cluster)

# remove codebook "k" from the memory
rm(k)


#### NZHS 2011/12 - adult ####

# R= refused, U= unable, P= pregnant, O=outliers
summary(HS11A$bmi) # 941 R, 637 U, 302 P, 20 O

HS11A$bmiscale <- as.numeric(as.character(HS11A$bmi))
summary(HS11A$bmiscale) # 1900 converted into NA

HS11A$gender <- as.factor(HS11A$gender)   # 0 = male, 1 = female

HS11A <- HS11A %>%
  mutate(bmic = ifelse(gender == 0 & age == 15 & bmiscale <16.98, 0,
                ifelse(gender == 0 & age == 16 & bmiscale <17.53, 0,
                ifelse(gender == 0 & age == 17 & bmiscale <18.04, 0,
                ifelse(gender == 1 & age == 15 & bmiscale <17.43, 0,
                ifelse(gender == 1 & age == 16 & bmiscale <17.9, 0,
                ifelse(gender == 1 & age == 17 & bmiscale <18.24, 0,
                ifelse(age >= 18 & bmiscale <18.5, 0,
                ifelse(gender == 0 & age == 15 & bmiscale >= 16.98 & bmiscale <23.28, 1,
                ifelse(gender == 0 & age == 16 & bmiscale >= 17.53 & bmiscale <23.89, 1,
                ifelse(gender == 0 & age == 17 & bmiscale >= 18.04 & bmiscale <24.46, 1,
                ifelse(gender == 1 & age == 15 & bmiscale >= 17.43 & bmiscale <23.89, 1,
                ifelse(gender == 1 & age == 16 & bmiscale >= 17.90 & bmiscale <24.34, 1,
```

```r
                    ifelse(gender == 1 & age == 17 & bmiscale >= 18.24 & bmiscale <24.70, 1,
                    ifelse(age >= 18 & bmiscale >= 18.5 & bmiscale <25, 1,
                    ifelse(gender == 0 & age == 15 & bmiscale >= 23.28 & bmiscale <28.32, 2,
                    ifelse(gender == 0 & age == 16 & bmiscale >= 23.89 & bmiscale <28.89, 2,
                    ifelse(gender == 0 & age == 17 & bmiscale >= 24.46 & bmiscale <29.43, 2,
                    ifelse(gender == 1 & age == 15 & bmiscale >= 23.89 & bmiscale <29.01, 2,
                    ifelse(gender == 1 & age == 16 & bmiscale >= 24.34 & bmiscale <29.40, 2,
                    ifelse(gender == 1 & age == 17 & bmiscale >= 24.70 & bmiscale <29.70, 2,
                    ifelse(age >= 18 & bmiscale >= 25 & bmiscale <30, 2,
                    ifelse(gender == 0 & age == 15 & bmiscale >=28.32, 3,
                    ifelse(gender == 0 & age == 16 & bmiscale >=28.89, 3,
                    ifelse(gender == 0 & age == 17 & bmiscale >=29.43, 3,
                    ifelse(gender == 1 & age == 15 & bmiscale >=29.01, 3,
                    ifelse(gender == 1 & age == 16 & bmiscale >=29.40, 3,
                    ifelse(gender == 1 & age == 17 & bmiscale >=29.70, 3,
                    ifelse(age >= 18 & bmiscale >= 30, 3 ,NA))))))))))))))))))))))))))))

HS11A$bmic <- as.factor(HS11A$bmic)
summary(HS11A$bmic)

# exclude underweight
a11 <- subset(HS11A, !bmic == 0 | is.na(bmic))


## Migration Status
# recode migrants <10-11 years into "Migrant", .K and .R into NA, and the rest into "Native"
a11$native <- recode(a11$A5_06, "2002"="Migrant", "2003"="Migrant", "2004"="Migrant", "2005"="
    Migrant",
                                  "2006"="Migrant", "2007"="Migrant", "2008"="Migrant", "2009"="
                                      Migrant",
                                  "2010"="Migrant", "2011"="Migrant", ".K"= "S", ".R"= "S",
                                  .default="Native")

summary(a11$native)
a11$native[a11$native == "S"] <- NA  #recode S into NA

a11$native <- factor(a11$native, levels=c("Native","Migrant"))

prop_frame(a11, "native")

```

```
## BMI
a11$bmic <- factor(a11$bmic, labels=c("Average","Overweight","Obese"))

prop_frame(a11, "bmic")

summary(a11$bmiscale)


## Sex
a11$gender <- factor(a11$gender, labels=c("Male","Female"))

prop_frame(a11, "gender")


## Ethnicity
# recode anyone ~ european
a11 = a11 %>%
  mutate(euro = ifelse(A5_03_01 == 1, 1, 0))

# recode anyone ~ maori
a11 = a11 %>%
  mutate(maori = ifelse(A5_03_02 == 1, 1, 0))

# recode anyone ~ pacific
a11 = a11 %>%
  mutate(pacific = ifelse(A5_03_03 == 1, 1,
                   ifelse(A5_03_04 == 1, 1,
                   ifelse(A5_03_05 == 1, 1,
                   ifelse(A5_03_06 == 1, 1, 0)))))

# recode anyone ~ asian
# starting from 2011, other asian will be coded as other ethnicity because there is no way to
    identify other asian
a11 = a11 %>%
  mutate(asian = ifelse(A5_03_07 == 1, 1,
                 ifelse(A5_03_08 == 1, 1, 0)))

# recode anyone ~ other
a11 = a11 %>%
```

```
  mutate(other = ifelse(A5_03_77 == 1, 1, 0))


a11 = a11 %>%
  mutate(eth_count = ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "Maori Only",
                     ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Pacific Only",
                     ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Asian Only",
                     ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 0, "European Only",
                     ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other",
                     ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+ Ethnicities (M)",
                     ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "2+ Ethnicities (M)",
                     ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Other",
                     ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Other",
                     ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other",
                     ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                     ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                     ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other",
                     ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other",
                     ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other",
                     ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+ Ethnicities (M)",
                     ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                     ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+ Ethnicities (M)",
```

```
                          ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other",
                          ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other",
                          ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other",
                          ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                          ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                          ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+ Ethnicities (M)",
                          ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other",
                          ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                          ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                          ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+ Ethnicities (M)",
                          ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other",
                          ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+ Ethnicities (M)",
                          ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+ Ethnicities (M)", NA)))))))))))))))))))))))))))))))

a11$eth_count = as.factor(a11$eth_count)

prop_frame(a11, "eth_count")


## Household Income
a11$hhinc <- ordered(a11$A5_24, levels=c("4","5","6","7","8","9","10","11","12",
                                              "13","14","15","16"))
sum(is.na(a11$hhinc))
sum(is.na(a11$hhinc))/length(a11$hhinc) # 369 NA, 9.5% missing

a11$hhinc <- as.numeric(as.character(a11$hhinc))
a11$hhinc <- cut(a11$hhinc, breaks=c(0,5,10,13,16))
```

```r
a11$hhinc <- factor(a11$hhinc, labels=c("<=$15,000","$15,001-$40,000","$40,001-$70,000",">$70,000"))

prop_frame(a11, "hhinc")


## Education
# 1= no qualification, 2= secondary school graduates, 3= tertiary school graduate
summary(a11$A5_14)
a11$secondary <- as.numeric(as.character(a11$A5_14))
a11$secondary <- recode(a11$secondary, "1" = "1", .default = "2") # 1= no degree, 2= secondary
    school grads
a11$secondary <- as.factor(a11$secondary)

summary(a11$A5_15)
a11$tertiary <- as.numeric(as.character(a11$A5_15))
a11$tertiary <- recode(a11$tertiary, "0" = "0", "1"="2",
                        "2"="2","3"="2","4"="2", .default = "3") #0= no tertiary, 2= secondary, 3=
                            tertiary
a11$tertiary <- as.factor(a11$tertiary)

a11 <- mutate(a11, edu = ifelse(secondary =="1" & tertiary =="0", 1,
                          ifelse(secondary == "2" & tertiary == "0", 2,
                          ifelse(tertiary == "2", 2,
                          ifelse(tertiary == "3", 3, 1)))))

a11$edu <- as.factor(a11$edu)
summary(a11$edu) #100 NA
sum(is.na(a11$edu))/length(a11$edu) # 0.95 % missing

a11$edu <- factor(a11$edu, labels=c("No Qualification","Secondary","Tertiary"))

prop_frame(a11, "edu")


## Deprivation Quintile
names(a11)[names(a11) == "nzdep06_quintile"] <- "dep"

prop_frame(a11, "dep")
```

```
## Fruit and Vegetable
summary(a11$fruit)
summary(a11$veges)
a11$fruit[a11$fruit == "K"] <- NA
a11$fruit[a11$fruit == "R"] <- NA
a11$veges[a11$veges == "K"] <- NA
a11$veges[a11$veges == "R"] <- NA
a11$fruit <- factor(a11$fruit, labels=c("No","Yes"))
a11$veges <- factor(a11$veges, labels=c("No","Yes"))

prop_frame(a11, "fruit")

prop_frame(a11, "veges")


## Alcohol Problem
a11$haz_drinker_all[a11$haz_drinker_all == "K" ] <- NA
a11$haz_drinker_all[a11$haz_drinker_all == "R" ] <- NA
a11$haz_drinker_all <- factor(a11$haz_drinker_all, labels = c("No Alcohol Problem","Alcohol Problem"
    ))

prop_frame(a11, "haz_drinker_all")


## Smoking Status
# checking for discrepancies
a11 %>%
  filter(!ex_smoker == "K" & !ex_smoker =="R") %>%
  group_by(ex_smoker) %>%
  summarise(n=n()) %>%
  mutate(prop=n/sum(n))

a11 %>%
  filter(!never_tried_smoking == "K" & !never_tried_smoking =="R") %>%
  group_by(never_tried_smoking) %>%
  summarise(n=n()) %>%
  mutate(prop=n/sum(n))

a11 %>%
  filter(!current_smoker == "K" & !current_smoker =="R") %>%
```

```r
  group_by(current_smoker) %>%
  summarise(n=n()) %>%
  mutate(prop=n/sum(n))

a11 %>%
  select(HHID, current_smoker, never_tried_smoking, ex_smoker) %>%
  filter(HHID == "1974")

# data number #1974 is incorrectly coded, adjustment is made for this data according to the answers from the questionnaires
a11 <- a11 %>%
  mutate(smoke = ifelse(HHID == "1974", "Ex Smoker",
                 ifelse(ex_smoker == 0 & current_smoker == 0, "Non Smoker",
                 ifelse(ex_smoker == 1 & current_smoker == 0, "Ex Smoker",
                        "Current Smoker"))))

a11$smoke <- factor(a11$smoke, levels= c("Non Smoker", "Ex Smoker","Current Smoker"))

prop_frame(a11, "smoke")


## Physical Activity
a11$active[a11$active =="K"] <- NA
a11$active[a11$active =="R"] <- NA

a11$active <- factor(a11$active, labels=c("Not Active", "Active"))

prop_frame(a11, "active")


## Sedentary Lifestyle
a11$sedentary[a11$sedentary =="K"] <- NA
a11$sedentary[a11$sedentary =="R"] <- NA

a11$sedentary <- factor(a11$sedentary, labels=c("Not Sedentary", "Sedentary"))

prop_frame(a11, "sedentary")


## Difficulty Climbing Several Steps of Stairs
```

```
summary(a11$A4_03)

a11$A4_03[a11$A4_03 == "."] <- NA
a11$A4_03[a11$A4_03 == ".K"] <- NA
a11$A4_03[a11$A4_03 == ".R"] <- NA
a11$stair <- factor(a11$A4_03, labels=c("A Lot Difficult","A Little Difficult","No Difficulty"))

prop_frame(a11, "stair")



#### NZHS 2011/12 - child ####

summary(HS11C$bmi) # 1037 Unobtainable, 415 Refused, 19 Outliers

HS11C$bmiscale <- as.numeric(as.character(HS11C$bmi))

HS11C$gender <- as.factor(HS11C$gender) #0: male, 1: female

HS11C <- HS11C %>%
  mutate(under =
  ifelse(gender == 0 & age == 2 & bmiscale <15.24, 0,
  ifelse(gender == 0 & age == 3 & bmiscale <14.83, 0,
  ifelse(gender == 0 & age == 4 & bmiscale <14.51, 0,
  ifelse(gender == 0 & age == 5 & bmiscale <14.26, 0,
  ifelse(gender == 0 & age == 6 & bmiscale <14.06, 0,
  ifelse(gender == 0 & age == 7 & bmiscale <14.00, 0,
  ifelse(gender == 0 & age == 8 & bmiscale <14.13, 0,
  ifelse(gender == 0 & age == 9 & bmiscale <14.36, 0,
  ifelse(gender == 0 & age == 10 & bmiscale <14.63, 0,
  ifelse(gender == 0 & age == 11 & bmiscale <14.96, 0,
  ifelse(gender == 0 & age == 12 & bmiscale <15.36, 0,
  ifelse(gender == 0 & age == 13 & bmiscale <15.84, 0,
  ifelse(gender == 0 & age == 14 & bmiscale <16.39, 0,
  ifelse(gender == 0 & age == 15 & bmiscale <16.98, 0,
  ifelse(gender == 0 & age == 16 & bmiscale <17.53, 0,
  ifelse(gender == 0 & age == 17 & bmiscale <18.04, 0,
  ifelse(gender == 1 & age == 2 & bmiscale <14.96, 0,
  ifelse(gender == 1 & age == 3 & bmiscale <14.60, 0,
  ifelse(gender == 1 & age == 4 & bmiscale <14.30, 0,
```

```
  ifelse(gender == 1 & age == 5 & bmiscale <14.04, 0,
  ifelse(gender == 1 & age == 6 & bmiscale <13.85, 0,
  ifelse(gender == 1 & age == 7 & bmiscale <13.83, 0,
  ifelse(gender == 1 & age == 8 & bmiscale <14.00, 0,
  ifelse(gender == 1 & age == 9 & bmiscale <14.26, 0,
  ifelse(gender == 1 & age == 10 & bmiscale <14.58, 0,
  ifelse(gender == 1 & age == 11 & bmiscale <15.03, 0,
  ifelse(gender == 1 & age == 12 & bmiscale <15.59, 0,
  ifelse(gender == 1 & age == 13 & bmiscale <16.23, 0,
  ifelse(gender == 1 & age == 14 & bmiscale <16.86, 0,
  ifelse(gender == 1 & age == 15 & bmiscale <17.43, 0,
  ifelse(gender == 1 & age == 16 & bmiscale <17.90, 0,
  ifelse(gender == 1 & age == 17 & bmiscale <18.24, 0,
  ifelse(age >= 18 & bmiscale <18.5, 0, NA))))))))))))))))))))))))))))))))

HS11C <- HS11C %>%
  mutate(average =
  ifelse(gender == 0 & age == 2 & bmiscale >=15.24 & bmiscale <18.36, 1,
  ifelse(gender == 0 & age == 3 & bmiscale >=14.83 & bmiscale <17.85, 1,
  ifelse(gender == 0 & age == 4 & bmiscale >=14.51 & bmiscale <17.52, 1,
  ifelse(gender == 0 & age == 5 & bmiscale >=14.26 & bmiscale <17.39, 1,
  ifelse(gender == 0 & age == 6 & bmiscale >=14.06 & bmiscale <17.52, 1,
  ifelse(gender == 0 & age == 7 & bmiscale >=14.00 & bmiscale <17.88, 1,
  ifelse(gender == 0 & age == 8 & bmiscale >=14.13 & bmiscale <18.41, 1,
  ifelse(gender == 0 & age == 9 & bmiscale >=14.36 & bmiscale <19.07, 1,
  ifelse(gender == 0 & age == 10 & bmiscale >=14.63 & bmiscale <19.80, 1,
  ifelse(gender == 0 & age == 11 & bmiscale >=14.96 & bmiscale <20.51, 1,
  ifelse(gender == 0 & age == 12 & bmiscale >=15.36 & bmiscale <21.20, 1,
  ifelse(gender == 0 & age == 13 & bmiscale >=15.84 & bmiscale <21.89, 1,
  ifelse(gender == 0 & age == 14 & bmiscale >=16.39 & bmiscale <22.60, 1,
  ifelse(gender == 0 & age == 15 & bmiscale >=16.98 & bmiscale <23.28, 1,
  ifelse(gender == 0 & age == 16 & bmiscale >=17.53 & bmiscale <23.89, 1,
  ifelse(gender == 0 & age == 17 & bmiscale >=18.04 & bmiscale <24.46, 1,
  ifelse(gender == 1 & age == 2 & bmiscale >=14.96 & bmiscale <18.09, 1,
  ifelse(gender == 1 & age == 3 & bmiscale >=14.60 & bmiscale <17.64, 1,
  ifelse(gender == 1 & age == 4 & bmiscale >=14.30 & bmiscale <17.35, 1,
  ifelse(gender == 1 & age == 5 & bmiscale >=14.04 & bmiscale <17.23, 1,
  ifelse(gender == 1 & age == 6 & bmiscale >=13.85 & bmiscale <17.33, 1,
  ifelse(gender == 1 & age == 7 & bmiscale >=13.83 & bmiscale <17.69, 1,
  ifelse(gender == 1 & age == 8 & bmiscale >=14.00 & bmiscale <18.28, 1,
```

```
  ifelse(gender == 1 & age == 9 & bmiscale >=14.26 & bmiscale <18.99, 1,
  ifelse(gender == 1 & age == 10 & bmiscale >=14.58 & bmiscale <19.78, 1,
  ifelse(gender == 1 & age == 11 & bmiscale >=15.03 & bmiscale <20.66, 1,
  ifelse(gender == 1 & age == 12 & bmiscale >=15.59 & bmiscale <21.59, 1,
  ifelse(gender == 1 & age == 13 & bmiscale >=16.23 & bmiscale <22.49, 1,
  ifelse(gender == 1 & age == 14 & bmiscale >=16.86 & bmiscale <23.27, 1,
  ifelse(gender == 1 & age == 15 & bmiscale >=17.43 & bmiscale <23.89, 1,
  ifelse(gender == 1 & age == 16 & bmiscale >=17.90 & bmiscale <24.34, 1,
  ifelse(gender == 1 & age == 17 & bmiscale >=18.24 & bmiscale <24.70, 1,
  ifelse(age >= 18 & bmiscale >=18.5 & bmiscale <25, 1, NA))))))))))))))))))))))))))))))))

HS11C <- HS11C %>%
  mutate(over=
  ifelse(gender == 0 & age == 2 & bmiscale >=18.36 & bmiscale <19.99, 2,
  ifelse(gender == 0 & age == 3 & bmiscale >=17.85 & bmiscale <19.50, 2,
  ifelse(gender == 0 & age == 4 & bmiscale >=17.52 & bmiscale <19.23, 2,
  ifelse(gender == 0 & age == 5 & bmiscale >=17.39 & bmiscale <19.27, 2,
  ifelse(gender == 0 & age == 6 & bmiscale >=17.52 & bmiscale <19.76, 2,
  ifelse(gender == 0 & age == 7 & bmiscale >=17.88 & bmiscale <20.59, 2,
  ifelse(gender == 0 & age == 8 & bmiscale >=18.41 & bmiscale <21.56, 2,
  ifelse(gender == 0 & age == 9 & bmiscale >=19.07 & bmiscale <22.71, 2,
  ifelse(gender == 0 & age == 10 & bmiscale >=19.80 & bmiscale <23.96, 2,
  ifelse(gender == 0 & age == 11 & bmiscale >=20.51 & bmiscale <25.07, 2,
  ifelse(gender == 0 & age == 12 & bmiscale >=21.20 & bmiscale <26.02, 2,
  ifelse(gender == 0 & age == 13 & bmiscale >=21.89 & bmiscale <26.87, 2,
  ifelse(gender == 0 & age == 14 & bmiscale >=22.60 & bmiscale <27.64, 2,
  ifelse(gender == 0 & age == 15 & bmiscale >=23.28 & bmiscale <28.32, 2,
  ifelse(gender == 0 & age == 16 & bmiscale >=23.89 & bmiscale <28.89, 2,
  ifelse(gender == 0 & age == 17 & bmiscale >=24.46 & bmiscale <29.43, 2,
  ifelse(gender == 1 & age == 2 & bmiscale >=18.09 & bmiscale <19.81, 2,
  ifelse(gender == 1 & age == 3 & bmiscale >=17.64 & bmiscale <19.38, 2,
  ifelse(gender == 1 & age == 4 & bmiscale >=17.35 & bmiscale <19.16, 2,
  ifelse(gender == 1 & age == 5 & bmiscale >=17.23 & bmiscale <19.20, 2,
  ifelse(gender == 1 & age == 6 & bmiscale >=17.33 & bmiscale <19.61, 2,
  ifelse(gender == 1 & age == 7 & bmiscale >=17.69 & bmiscale <20.39, 2,
  ifelse(gender == 1 & age == 8 & bmiscale >=18.28 & bmiscale <21.44, 2,
  ifelse(gender == 1 & age == 9 & bmiscale >=18.99 & bmiscale <22.66, 2,
  ifelse(gender == 1 & age == 10 & bmiscale >=19.78 & bmiscale <23.97, 2,
  ifelse(gender == 1 & age == 11 & bmiscale >=20.66 & bmiscale <25.25, 2,
  ifelse(gender == 1 & age == 12 & bmiscale >=21.59 & bmiscale <26.47, 2,
```

```
  ifelse(gender == 1 & age == 13 & bmiscale >=22.49 & bmiscale <27.57, 2,
  ifelse(gender == 1 & age == 14 & bmiscale >=23.27 & bmiscale <28.42, 2,
  ifelse(gender == 1 & age == 15 & bmiscale >=23.89 & bmiscale <29.01, 2,
  ifelse(gender == 1 & age == 16 & bmiscale >=24.34 & bmiscale <29.40, 2,
  ifelse(gender == 1 & age == 17 & bmiscale >=24.70 & bmiscale <29.70, 2,
  ifelse(age >= 18 & bmiscale >=25 & bmiscale <30, 2, NA))))))))))))))))))))))))))))))))

HS11C <- HS11C %>%
  mutate(obes=
  ifelse(gender == 0 & age == 2 & bmiscale >=19.99, 3,
  ifelse(gender == 0 & age == 3 & bmiscale >=19.50, 3,
  ifelse(gender == 0 & age == 4 & bmiscale >=19.23, 3,
  ifelse(gender == 0 & age == 5 & bmiscale >=19.27, 3,
  ifelse(gender == 0 & age == 6 & bmiscale >=19.76, 3,
  ifelse(gender == 0 & age == 7 & bmiscale >=20.59, 3,
  ifelse(gender == 0 & age == 8 & bmiscale >=21.56, 3,
  ifelse(gender == 0 & age == 9 & bmiscale >=22.71, 3,
  ifelse(gender == 0 & age == 10 & bmiscale >=23.96, 3,
  ifelse(gender == 0 & age == 11 & bmiscale >=25.07, 3,
  ifelse(gender == 0 & age == 12 & bmiscale >=26.02, 3,
  ifelse(gender == 0 & age == 13 & bmiscale >=26.87, 3,
  ifelse(gender == 0 & age == 14 & bmiscale >=27.64, 3,
  ifelse(gender == 0 & age == 15 & bmiscale >=28.32, 3,
  ifelse(gender == 0 & age == 16 & bmiscale >=28.89, 3,
  ifelse(gender == 0 & age == 17 & bmiscale >=29.43, 3,
  ifelse(gender == 1 & age == 2 & bmiscale >=19.81, 3,
  ifelse(gender == 1 & age == 3 & bmiscale >=19.38, 3,
  ifelse(gender == 1 & age == 4 & bmiscale >=19.16, 3,
  ifelse(gender == 1 & age == 5 & bmiscale >=19.20, 3,
  ifelse(gender == 1 & age == 6 & bmiscale >=19.61, 3,
  ifelse(gender == 1 & age == 7 & bmiscale >=20.39, 3,
  ifelse(gender == 1 & age == 8 & bmiscale >=21.44, 3,
  ifelse(gender == 1 & age == 9 & bmiscale >=22.66, 3,
  ifelse(gender == 1 & age == 10 & bmiscale >=23.97, 3,
  ifelse(gender == 1 & age == 11 & bmiscale >=25.25, 3,
  ifelse(gender == 1 & age == 12 & bmiscale >=26.47, 3,
  ifelse(gender == 1 & age == 13 & bmiscale >=27.57, 3,
  ifelse(gender == 1 & age == 14 & bmiscale >=28.42, 3,
  ifelse(gender == 1 & age == 15 & bmiscale >=29.01, 3,
  ifelse(gender == 1 & age == 16 & bmiscale >=29.40, 3,
```

```r
  ifelse(gender == 1 & age == 17 & bmiscale >=29.70, 3,
  ifelse(age >= 18 & bmiscale >=30, 3, NA))))))))))))))))))))))))))))))))


# 0: underweight, 1: average, 2: overweight, 3: obese
HS11C <- HS11C %>%
  mutate(bmic = ifelse(under == 0 & is.na(average) & is.na(over) & is.na(obes), 0,
                ifelse(average == 1 & is.na(under) & is.na(over) & is.na(obes), 1,
                ifelse(over == 2 & is.na(under) & is.na(average) & is.na(obes), 2,
                ifelse(obes == 3 & is.na(under) & is.na(over) & is.na(average), 3, NA)))))

HS11C$bmic <- as.factor(HS11C$bmic)


# excluded underweight in the bmi data
c11 <- subset(HS11C, !bmic == 0 | is.na(bmic))


## BMI
c11$bmic <- factor(c11$bmic, labels=c("Average","Overweight","Obese"))

prop_frame(c11, "bmic")


## Sex
c11$gender <- factor(c11$gender, labels=c("Male","Female"))

prop_frame(c11, "gender")


## Age
summary(c11$age)


## Ethnicity
# recode anyone ~ european
c11 = c11 %>%
  mutate(euro = ifelse(C4_03_01 == 1, 1, 0))

# recode anyone ~ maori
```

```
c11 = c11 %>%
  mutate(maori = ifelse(C4_03_02 == 1, 1, 0))

# recode anyone ~ pacific
c11 = c11 %>%
  mutate(pacific = ifelse(C4_03_03 == 1, 1,
                   ifelse(C4_03_04 == 1, 1,
                   ifelse(C4_03_05 == 1, 1,
                   ifelse(C4_03_06 == 1, 1, 0)))))

# recode anyone ~ asian (Indian and Chinese)
c11 = c11 %>%
  mutate(asian = ifelse(C4_03_07 == 1, 1,
                 ifelse(C4_03_08 == 1, 1, 0)))

# recode anyone ~ other
c11 = c11 %>%
  mutate(other = ifelse(C4_03_77 == 1, 1, 0))


c11 = c11 %>%
  mutate(eth_count = ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "Maori Only",
                     ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Pacific Only",
                     ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Asian Only",
                     ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 0, "European Only",
                     ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other",
                     ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+ Ethnicities (M)",
                     ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "2+ Ethnicities (M)",
                     ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Other",
                     ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Other",
```

```r
            ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other",
            ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+ Ethnicities (M)",
            ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+ Ethnicities (M)",
            ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other",
            ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other",
            ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other",
            ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+ Ethnicities (M)",
            ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+ Ethnicities (M)",
            ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+ Ethnicities (M)",
            ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other",
            ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other",
            ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other",
            ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+ Ethnicities (M)",
            ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+ Ethnicities (M)",
            ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+ Ethnicities (M)",
            ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other",
            ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+ Ethnicities (M)",
            ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+ Ethnicities (M)",
            ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+ Ethnicities (M)",
            ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other",
```

```
                    ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
                        Ethnicities (M)",
                    ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
                        Ethnicities (M)", NA))))))))))))))))))))))))))))))))

c11$eth_count <- as.factor(c11$eth_count)

prop_frame(c11, "eth_count")


## Household Income
# merge adult 2011 with child 2011 data by HHID to match income and education data from the same
    household
ac11 <- left_join(select(c11, HHID), select(a11, HHID, A5_24, A5_14, A5_15), by="HHID")

summary(ac11$A5_24)

ac11$hhinc <- as.numeric(as.character(ac11$A5_24))
ac11$hhinc <- cut(ac11$hhinc, breaks=c(0,5,10,13,16))
ac11$hhinc <- factor(ac11$hhinc, labels=c("<=$15,000","$15,001-$40,000","$40,001-$70,000",">$70,000"
    ))

# return the value to the original dataset
c11$hhinc <- ac11$hhinc

prop_frame(c11, "hhinc")


## Education
ac11$secondary <- as.numeric(as.character(ac11$A5_14))
ac11$secondary <- recode(ac11$secondary, "1" = "1", .default = "2") # 1= no degree, 2= secondary
    school grads
ac11$secondary <- as.factor(ac11$secondary)

ac11$tertiary <- as.numeric(as.character(ac11$A5_15))
ac11$tertiary <- recode(ac11$tertiary, "0" = "0", "1"="2",
                      "2"="2","3"="2","4"="2", .default = "3") #0 = no tertiary, 2= secondary, 3=
                          tertiary
ac11$tertiary <- as.factor(ac11$tertiary)

```

```r
# prioritise the highes educational qualification
ac11 <- mutate(ac11, edu = ifelse(secondary =="1" & tertiary =="0", 1,
                          ifelse(secondary == "2" & tertiary == "0", 2,
                          ifelse(tertiary =="2", 2,
                          ifelse(tertiary == "3", 3, 1)))))

# return the information to original data set
c11$edu <- ac11$edu
c11$edu <- as.factor(c11$edu)
c11$edu <- factor(c11$edu, labels=c("No Qualification","Secondary","Tertiary"))

prop_frame(c11, "edu")

# remove ac11 from the memory
rm(ac11)

# education missing value
sum(is.na(c11$edu))
sum(is.na(c11$edu))/length(c11$edu) # 455 NA (15.27%)


## Deprivation Quintile
names(c11)[names(c11) == "nzdep06_quintile"] <- "dep"

prop_frame(c11, "dep")


## Fruit and Vegetable
c11$veges <- c11$kidveges_new
c11$fruit <- c11$kidfruit_2serves

c11$fruit[c11$fruit == "K"] <- NA
c11$fruit[c11$fruit == "X"] <- NA
c11$veges[c11$veges == "K"] <- NA
c11$veges[c11$veges == "R"] <- NA
c11$veges[c11$veges == "X"] <- NA
c11$fruit <- factor(c11$fruit, labels=c("No","Yes"))
c11$veges <- factor(c11$veges, labels=c("No","Yes"))

prop_frame(c11, "fruit")
```

```
prop_frame(c11, "veges")


## Soft Drink and Fast Food
summary(c11$C3_09) #14 NA, 18 K
c11$C3_09 <-as.numeric(as.character(c11$C3_09))

c11$softd <- cut(c11$C3_09, c(-Inf,1,2,4,Inf), right=FALSE)
summary(c11$softd)

c11$softd <- factor(c11$softd, labels=c("0/week","1/week","2-3/week","4+/week"))

prop_frame(c11, "softd")

#fast food
summary(c11$C3_10) # 14 NA, 8 K
c11$C3_10 <- as.numeric(as.character(c11$C3_10))

c11$fastf <- cut(c11$C3_10, c(0,1,2,4,Inf), right=FALSE)

c11$fastf <- factor(c11$fastf, labels=c("0/week","1/week","2-3/week","4+/week"))

prop_frame(c11, "fastf")



#### NZHS 2012/13 - adult ####

summary(HS12A$bmi) # 427 R, 423 U, 214 P, 4 O

HS12A$bmiscale <- as.numeric(as.character(HS12A$bmi)) # force non-numeric into NA

HS12A$gender <- as.factor(HS12A$gender)

HS12A <- HS12A %>%
  mutate(bmic = ifelse(gender == 0 & age == 15 & bmiscale <16.98, 0,
                 ifelse(gender == 0 & age == 16 & bmiscale <17.53, 0,
                 ifelse(gender == 0 & age == 17 & bmiscale <18.04, 0,
                 ifelse(gender == 1 & age == 15 & bmiscale <17.43, 0,
                 ifelse(gender == 1 & age == 16 & bmiscale <17.9, 0,
```

```
                    ifelse(gender == 1 & age == 17 & bmiscale <18.24, 0,
                    ifelse(age >= 18 & bmiscale <18.5, 0,
                    ifelse(gender == 0 & age == 15 & bmiscale >= 16.98 & bmiscale <23.28, 1,
                    ifelse(gender == 0 & age == 16 & bmiscale >= 17.53 & bmiscale <23.89, 1,
                    ifelse(gender == 0 & age == 17 & bmiscale >= 18.04 & bmiscale <24.46, 1,
                    ifelse(gender == 1 & age == 15 & bmiscale >= 17.43 & bmiscale <23.89, 1,
                    ifelse(gender == 1 & age == 16 & bmiscale >= 17.90 & bmiscale <24.34, 1,
                    ifelse(gender == 1 & age == 17 & bmiscale >= 18.24 & bmiscale <24.70, 1,
                    ifelse(age >= 18 & bmiscale >= 18.5 & bmiscale <25, 1,
                    ifelse(gender == 0 & age == 15 & bmiscale >= 23.28 & bmiscale <28.32, 2,
                    ifelse(gender == 0 & age == 16 & bmiscale >= 23.89 & bmiscale <28.89, 2,
                    ifelse(gender == 0 & age == 17 & bmiscale >= 24.46 & bmiscale <29.43, 2,
                    ifelse(gender == 1 & age == 15 & bmiscale >= 23.89 & bmiscale <29.01, 2,
                    ifelse(gender == 1 & age == 16 & bmiscale >= 24.34 & bmiscale <29.40, 2,
                    ifelse(gender == 1 & age == 17 & bmiscale >= 24.70 & bmiscale <29.70, 2,
                    ifelse(age >= 18 & bmiscale >= 25 & bmiscale <30, 2,
                    ifelse(gender == 0 & age == 15 & bmiscale >=28.32, 3,
                    ifelse(gender == 0 & age == 16 & bmiscale >=28.89, 3,
                    ifelse(gender == 0 & age == 17 & bmiscale >=29.43, 3,
                    ifelse(gender == 1 & age == 15 & bmiscale >=29.01, 3,
                    ifelse(gender == 1 & age == 16 & bmiscale >=29.40, 3,
                    ifelse(gender == 1 & age == 17 & bmiscale >=29.70, 3,
                    ifelse(age >= 18 & bmiscale >= 30, 3 ,NA))))))))))))))))))))))))))))

HS12A$bmic <- as.factor(HS12A$bmic)

# exclude underweight
a12 <- subset(HS12A, !bmic == 0 | is.na(bmic))


## Migration Status
# recode people who lived in NZ for less than 10-11 years as Migrants
a12$native <- recode(a12$A5_06, "2003"="Migrant", "2004"="Migrant", "2005"="Migrant",
                                    "2006"="Migrant", "2007"="Migrant", "2008"="Migrant", "2009"="
                                        Migrant",
                                    "2010"="Migrant", "2011"="Migrant", "2012"="Migrant", ".K"= "S", ".R
                                        "= "S",
                                    .default="Native")
a12$native[a12$native == "S"] <- NA  #recode S into NA
a12$native <- droplevels(a12$native)
```

```r

prop_frame(a12, "native")


## BMI
a12$bmic <- factor(a12$bmic, labels =c("Average","Overweight","Obese"))

prop_frame(a12, "bmic")


## Sex
a12$gender <- factor(a12$gender, labels=c("Male","Female"))

prop_frame(a12, "gender")


## Age
summary(a12$age)


## Ethnicity
# recode anyone ~ european
a12 = a12 %>%
  mutate(euro = ifelse(A5_03_01 == 1, 1, 0))

# recode anyone ~ maori
a12 = a12 %>%
  mutate(maori = ifelse(A5_03_02 == 1, 1, 0))

# recode anyone ~ pacific
a12 = a12 %>%
  mutate(pacific = ifelse(A5_03_03 == 1, 1,
                   ifelse(A5_03_04 == 1, 1,
                   ifelse(A5_03_05 == 1, 1,
                   ifelse(A5_03_06 == 1, 1, 0)))))

# recode anyone ~ asian
a12 = a12 %>%
  mutate(asian = ifelse(A5_03_07 == 1, 1,
                 ifelse(A5_03_08 == 1, 1, 0)))
```

```
# recode anyone ~ other
a12 = a12 %>%
  mutate(other = ifelse(A5_03_77 == 1, 1, 0))


a12 = a12 %>%
  mutate(eth_count = ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "Maori Only",
                    ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Pacific Only",
                    ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Asian Only",
                    ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 0, "European Only",
                    ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other",
                    ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+ Ethnicities (M)",
                    ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "2+ Ethnicities (M)",
                    ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Other",
                    ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Other",
                    ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other",
                    ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                    ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                    ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other",
                    ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other",
                    ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other",
                    ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+ Ethnicities (M)",
```

```r
                    ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+
                        Ethnicities (M)",
                    ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+
                        Ethnicities (M)",
                    ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other"
                        ,
                    ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other"
                        ,
                    ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other"
                        ,
                    ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+
                        Ethnicities (M)",
                    ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+
                        Ethnicities (M)",
                    ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+
                        Ethnicities (M)",
                    ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other"
                        ,
                    ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+
                        Ethnicities (M)",
                    ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+
                        Ethnicities (M)",
                    ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+
                        Ethnicities (M)",
                    ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other"
                        ,
                    ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
                        Ethnicities (M)",
                    ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
                        Ethnicities (M)", NA))))))))))))))))))))))))))))))

a12$eth_count <- as.factor(a12$eth_count)

prop_frame(a12, "eth_count")


## Household Income
a12$hhinc <- ordered(a12$A5_24, levels=c("4","5","6","7","8","9","10","11","12",
                                       "13","14","15","16"))

```

```
a12$hhinc <- as.numeric(as.character(a12$hhinc))
a12$hhinc <- cut(a12$hhinc, breaks=c(0,5,10,13,16))
a12$hhinc <- factor(a12$hhinc, labels=c("<=$15,000","$15,001-$40,000","$40,001-$70,000",">$70,000"))

prop_frame(a12, "hhinc")


## Education
a12$secondary <- as.numeric(as.character(a12$A5_14))
a12$secondary <- recode(a12$secondary, "1" = "1", .default = "2") # 1= no degree, 2= secondary
    school grads
a12$secondary <- as.factor(a12$secondary)

a12$tertiary <- as.numeric(as.character(a12$A5_15)) # force .K and R. into NA
a12$tertiary <- recode(a12$tertiary, "0" = "0","1"="2","2"="2",
                        "3"="2", "4"="2", .default = "3") #0 = no tertiary, 2= secondary, 3= tertiary
a12$tertiary <- as.factor(a12$tertiary)

a12 <- mutate(a12, edu = ifelse(secondary =="1" & tertiary =="0", 1,
                          ifelse(secondary == "2" & tertiary == "0", 2,
                          ifelse(tertiary == "2", 2,
                          ifelse(tertiary == "3", 3, 1)))))

a12$edu <- as.factor(a12$edu)

a12$edu <- factor(a12$edu, labels=c("No Qualification","Secondary","Tertiary"))

prop_frame(a12, "edu")


## Deprivation Quintile
names(a12)[names(a12) == "nzdep06_quintile"] <- "dep"

prop_frame(a12, "dep")


## Fruit and Vegetable
summary(a12$fruit) # 10 K, 1 R
a12$fruit[a12$fruit =="K"] <- NA
a12$fruit[a12$fruit =="R"] <- NA
```

```
a12$fruit <- factor(a12$fruit , labels=c("No","Yes"))

prop_frame(a12, "fruit")

summary(a12$veges) # 7 K, 1 R
a12$veges[a12$veges =="K"] <- NA
a12$veges[a12$veges =="R"] <- NA

a12$veges <- factor(a12$veges, labels=c("No","Yes"))

prop_frame(a12, "veges")


## Alcohol Problom
summary(a12$haz_drinker_all) # 130 K, 14 R

a12$haz_drinker_all[a12$haz_drinker_all == "K"] <- NA
a12$haz_drinker_all[a12$haz_drinker_all == "R"] <- NA

a12$haz_drinker_all <- factor(a12$haz_drinker_all , labels=c("No Alcohol Problem",
                                                             "Alcohol Problem"))

prop_frame(a12, "haz_drinker_all")


## Smoking Status
a12 <- a12 %>%
  mutate(smoke = ifelse(current_smoker == 1 & ex_smoker == 0, 3,
                 ifelse(current_smoker == 0 & ex_smoker == 1, 2,
                  ifelse(current_smoker == "K", NA,
                  ifelse(current_smoker == "R", NA,
                  ifelse(ex_smoker =="K", NA,
                  ifelse(ex_smoker =="R", NA, 1)))))))

a12$smoke <- as.factor(a12$smoke)
a12$smoke <- factor(a12$smoke, labels=c("Non Smoker","Ex Smoker","Current Smoker"))

prop_frame(a12, "smoke")

```

```
## Physically Active
summary(a12$active) # 131 K, 3 R
a12$active[a12$active == "K"] <- NA
a12$active[a12$active == "R"] <- NA

a12$active <- factor(a12$active, labels=c("Not Active","Active"))

summary(a12$sedentary) # 129 K, 2 R

a12$sedentary[a12$sedentary == "K"] <- NA
a12$sedentary[a12$sedentary == "R"] <- NA

a12$sedentary <- factor(a12$sedentary, labels=c("Not Sedentary", "Sedentary"))

prop_frame(a12, "active")
prop_frame(a12, "sedentary")


## Difficulty Climbing Several Steps of Stairs
summary(a12$A4_03)
a12$A4_03[a12$A4_03 == ".K"] <- NA
a12$A4_03[a12$A4_03 == ".R"] <- NA

a12$stair <- factor(a12$A4_03, labels=c("A Lot Difficult","A Little Difficult",
                                        "No Difficulty"))

prop_frame(a12, "stair")



#### NZHS 2012/13 - child ####

### SUBSETTING ###
HS12C$bmiscale <- as.numeric(as.character(HS12C$bmi))

HS12C$gender <- as.factor(HS12C$gender)

HS12C <- HS12C %>%
  mutate(under =
```

```r
  ifelse(gender == 0 & age == 2 & bmiscale <15.24, 0,
  ifelse(gender == 0 & age == 3 & bmiscale <14.83, 0,
  ifelse(gender == 0 & age == 4 & bmiscale <14.51, 0,
  ifelse(gender == 0 & age == 5 & bmiscale <14.26, 0,
  ifelse(gender == 0 & age == 6 & bmiscale <14.06, 0,
  ifelse(gender == 0 & age == 7 & bmiscale <14.00, 0,
  ifelse(gender == 0 & age == 8 & bmiscale <14.13, 0,
  ifelse(gender == 0 & age == 9 & bmiscale <14.36, 0,
  ifelse(gender == 0 & age == 10 & bmiscale <14.63, 0,
  ifelse(gender == 0 & age == 11 & bmiscale <14.96, 0,
  ifelse(gender == 0 & age == 12 & bmiscale <15.36, 0,
  ifelse(gender == 0 & age == 13 & bmiscale <15.84, 0,
  ifelse(gender == 0 & age == 14 & bmiscale <16.39, 0,
  ifelse(gender == 0 & age == 15 & bmiscale <16.98, 0,
  ifelse(gender == 0 & age == 16 & bmiscale <17.53, 0,
  ifelse(gender == 0 & age == 17 & bmiscale <18.04, 0,
  ifelse(gender == 1 & age == 2 & bmiscale <14.96, 0,
  ifelse(gender == 1 & age == 3 & bmiscale <14.60, 0,
  ifelse(gender == 1 & age == 4 & bmiscale <14.30, 0,
  ifelse(gender == 1 & age == 5 & bmiscale <14.04, 0,
  ifelse(gender == 1 & age == 6 & bmiscale <13.85, 0,
  ifelse(gender == 1 & age == 7 & bmiscale <13.83, 0,
  ifelse(gender == 1 & age == 8 & bmiscale <14.00, 0,
  ifelse(gender == 1 & age == 9 & bmiscale <14.26, 0,
  ifelse(gender == 1 & age == 10 & bmiscale <14.58, 0,
  ifelse(gender == 1 & age == 11 & bmiscale <15.03, 0,
  ifelse(gender == 1 & age == 12 & bmiscale <15.59, 0,
  ifelse(gender == 1 & age == 13 & bmiscale <16.23, 0,
  ifelse(gender == 1 & age == 14 & bmiscale <16.86, 0,
  ifelse(gender == 1 & age == 15 & bmiscale <17.43, 0,
  ifelse(gender == 1 & age == 16 & bmiscale <17.90, 0,
  ifelse(gender == 1 & age == 17 & bmiscale <18.24, 0,
  ifelse(age >= 18 & bmiscale <18.5, 0, NA)))))))))))))))))))))))))))))))))

HS12C <- HS12C %>%
  mutate(average =
  ifelse(gender == 0 & age == 2 & bmiscale >=15.24 & bmiscale <18.36, 1,
  ifelse(gender == 0 & age == 3 & bmiscale >=14.83 & bmiscale <17.85, 1,
  ifelse(gender == 0 & age == 4 & bmiscale >=14.51 & bmiscale <17.52, 1,
  ifelse(gender == 0 & age == 5 & bmiscale >=14.26 & bmiscale <17.39, 1,
```

```
  ifelse(gender == 0 & age == 6 & bmiscale >=14.06 & bmiscale <17.52, 1,
  ifelse(gender == 0 & age == 7 & bmiscale >=14.00 & bmiscale <17.88, 1,
  ifelse(gender == 0 & age == 8 & bmiscale >=14.13 & bmiscale <18.41, 1,
  ifelse(gender == 0 & age == 9 & bmiscale >=14.36 & bmiscale <19.07, 1,
  ifelse(gender == 0 & age == 10 & bmiscale >=14.63 & bmiscale <19.80, 1,
  ifelse(gender == 0 & age == 11 & bmiscale >=14.96 & bmiscale <20.51, 1,
  ifelse(gender == 0 & age == 12 & bmiscale >=15.36 & bmiscale <21.20, 1,
  ifelse(gender == 0 & age == 13 & bmiscale >=15.84 & bmiscale <21.89, 1,
  ifelse(gender == 0 & age == 14 & bmiscale >=16.39 & bmiscale <22.60, 1,
  ifelse(gender == 0 & age == 15 & bmiscale >=16.98 & bmiscale <23.28, 1,
  ifelse(gender == 0 & age == 16 & bmiscale >=17.53 & bmiscale <23.89, 1,
  ifelse(gender == 0 & age == 17 & bmiscale >=18.04 & bmiscale <24.46, 1,
  ifelse(gender == 1 & age == 2 & bmiscale >=14.96 & bmiscale <18.09, 1,
  ifelse(gender == 1 & age == 3 & bmiscale >=14.60 & bmiscale <17.64, 1,
  ifelse(gender == 1 & age == 4 & bmiscale >=14.30 & bmiscale <17.35, 1,
  ifelse(gender == 1 & age == 5 & bmiscale >=14.04 & bmiscale <17.23, 1,
  ifelse(gender == 1 & age == 6 & bmiscale >=13.85 & bmiscale <17.33, 1,
  ifelse(gender == 1 & age == 7 & bmiscale >=13.83 & bmiscale <17.69, 1,
  ifelse(gender == 1 & age == 8 & bmiscale >=14.00 & bmiscale <18.28, 1,
  ifelse(gender == 1 & age == 9 & bmiscale >=14.26 & bmiscale <18.99, 1,
  ifelse(gender == 1 & age == 10 & bmiscale >=14.58 & bmiscale <19.78, 1,
  ifelse(gender == 1 & age == 11 & bmiscale >=15.03 & bmiscale <20.66, 1,
  ifelse(gender == 1 & age == 12 & bmiscale >=15.59 & bmiscale <21.59, 1,
  ifelse(gender == 1 & age == 13 & bmiscale >=16.23 & bmiscale <22.49, 1,
  ifelse(gender == 1 & age == 14 & bmiscale >=16.86 & bmiscale <23.27, 1,
  ifelse(gender == 1 & age == 15 & bmiscale >=17.43 & bmiscale <23.89, 1,
  ifelse(gender == 1 & age == 16 & bmiscale >=17.90 & bmiscale <24.34, 1,
  ifelse(gender == 1 & age == 17 & bmiscale >=18.24 & bmiscale <24.70, 1,
  ifelse(age >= 18 & bmiscale >=18.5 & bmiscale <25, 1, NA)))))))))))))))))))))))))))))))))

HS12C <- HS12C %>%
  mutate(over=
  ifelse(gender == 0 & age == 2 & bmiscale >=18.36 & bmiscale <19.99, 2,
  ifelse(gender == 0 & age == 3 & bmiscale >=17.85 & bmiscale <19.50, 2,
  ifelse(gender == 0 & age == 4 & bmiscale >=17.52 & bmiscale <19.23, 2,
  ifelse(gender == 0 & age == 5 & bmiscale >=17.39 & bmiscale <19.27, 2,
  ifelse(gender == 0 & age == 6 & bmiscale >=17.52 & bmiscale <19.76, 2,
  ifelse(gender == 0 & age == 7 & bmiscale >=17.88 & bmiscale <20.59, 2,
  ifelse(gender == 0 & age == 8 & bmiscale >=18.41 & bmiscale <21.56, 2,
  ifelse(gender == 0 & age == 9 & bmiscale >=19.07 & bmiscale <22.71, 2,
```

```r
  ifelse(gender == 0 & age == 10 & bmiscale >=19.80 & bmiscale <23.96, 2,
  ifelse(gender == 0 & age == 11 & bmiscale >=20.51 & bmiscale <25.07, 2,
  ifelse(gender == 0 & age == 12 & bmiscale >=21.20 & bmiscale <26.02, 2,
  ifelse(gender == 0 & age == 13 & bmiscale >=21.89 & bmiscale <26.87, 2,
  ifelse(gender == 0 & age == 14 & bmiscale >=22.60 & bmiscale <27.64, 2,
  ifelse(gender == 0 & age == 15 & bmiscale >=23.28 & bmiscale <28.32, 2,
  ifelse(gender == 0 & age == 16 & bmiscale >=23.89 & bmiscale <28.89, 2,
  ifelse(gender == 0 & age == 17 & bmiscale >=24.46 & bmiscale <29.43, 2,
  ifelse(gender == 1 & age == 2 & bmiscale >=18.09 & bmiscale <19.81, 2,
  ifelse(gender == 1 & age == 3 & bmiscale >=17.64 & bmiscale <19.38, 2,
  ifelse(gender == 1 & age == 4 & bmiscale >=17.35 & bmiscale <19.16, 2,
  ifelse(gender == 1 & age == 5 & bmiscale >=17.23 & bmiscale <19.20, 2,
  ifelse(gender == 1 & age == 6 & bmiscale >=17.33 & bmiscale <19.61, 2,
  ifelse(gender == 1 & age == 7 & bmiscale >=17.69 & bmiscale <20.39, 2,
  ifelse(gender == 1 & age == 8 & bmiscale >=18.28 & bmiscale <21.44, 2,
  ifelse(gender == 1 & age == 9 & bmiscale >=18.99 & bmiscale <22.66, 2,
  ifelse(gender == 1 & age == 10 & bmiscale >=19.78 & bmiscale <23.97, 2,
  ifelse(gender == 1 & age == 11 & bmiscale >=20.66 & bmiscale <25.25, 2,
  ifelse(gender == 1 & age == 12 & bmiscale >=21.59 & bmiscale <26.47, 2,
  ifelse(gender == 1 & age == 13 & bmiscale >=22.49 & bmiscale <27.57, 2,
  ifelse(gender == 1 & age == 14 & bmiscale >=23.27 & bmiscale <28.42, 2,
  ifelse(gender == 1 & age == 15 & bmiscale >=23.89 & bmiscale <29.01, 2,
  ifelse(gender == 1 & age == 16 & bmiscale >=24.34 & bmiscale <29.40, 2,
  ifelse(gender == 1 & age == 17 & bmiscale >=24.70 & bmiscale <29.70, 2,
  ifelse(age >= 18 & bmiscale >=25 & bmiscale <30, 2, NA))))))))))))))))))))))))))))))))))

HS12C <- HS12C %>%
  mutate(obes=
  ifelse(gender == 0 & age == 2 & bmiscale >=19.99, 3,
  ifelse(gender == 0 & age == 3 & bmiscale >=19.50, 3,
  ifelse(gender == 0 & age == 4 & bmiscale >=19.23, 3,
  ifelse(gender == 0 & age == 5 & bmiscale >=19.27, 3,
  ifelse(gender == 0 & age == 6 & bmiscale >=19.76, 3,
  ifelse(gender == 0 & age == 7 & bmiscale >=20.59, 3,
  ifelse(gender == 0 & age == 8 & bmiscale >=21.56, 3,
  ifelse(gender == 0 & age == 9 & bmiscale >=22.71, 3,
  ifelse(gender == 0 & age == 10 & bmiscale >=23.96, 3,
  ifelse(gender == 0 & age == 11 & bmiscale >=25.07, 3,
  ifelse(gender == 0 & age == 12 & bmiscale >=26.02, 3,
  ifelse(gender == 0 & age == 13 & bmiscale >=26.87, 3,
```

```r
  ifelse(gender == 0 & age == 14 & bmiscale >=27.64, 3,
  ifelse(gender == 0 & age == 15 & bmiscale >=28.32, 3,
  ifelse(gender == 0 & age == 16 & bmiscale >=28.89, 3,
  ifelse(gender == 0 & age == 17 & bmiscale >=29.43, 3,
  ifelse(gender == 1 & age == 2 & bmiscale >=19.81, 3,
  ifelse(gender == 1 & age == 3 & bmiscale >=19.38, 3,
  ifelse(gender == 1 & age == 4 & bmiscale >=19.16, 3,
  ifelse(gender == 1 & age == 5 & bmiscale >=19.20, 3,
  ifelse(gender == 1 & age == 6 & bmiscale >=19.61, 3,
  ifelse(gender == 1 & age == 7 & bmiscale >=20.39, 3,
  ifelse(gender == 1 & age == 8 & bmiscale >=21.44, 3,
  ifelse(gender == 1 & age == 9 & bmiscale >=22.66, 3,
  ifelse(gender == 1 & age == 10 & bmiscale >=23.97, 3,
  ifelse(gender == 1 & age == 11 & bmiscale >=25.25, 3,
  ifelse(gender == 1 & age == 12 & bmiscale >=26.47, 3,
  ifelse(gender == 1 & age == 13 & bmiscale >=27.57, 3,
  ifelse(gender == 1 & age == 14 & bmiscale >=28.42, 3,
  ifelse(gender == 1 & age == 15 & bmiscale >=29.01, 3,
  ifelse(gender == 1 & age == 16 & bmiscale >=29.40, 3,
  ifelse(gender == 1 & age == 17 & bmiscale >=29.70, 3,
  ifelse(age >= 18 & bmiscale >=30, 3, NA))))))))))))))))))))))))))))))))))

# 0: underweight, 1: average, 2: overweight, 3: obese
HS12C <- HS12C %>%
  mutate(bmic = ifelse(under == 0 & is.na(average) & is.na(over) & is.na(obes), 0,
                ifelse(average == 1 & is.na(under) & is.na(over) & is.na(obes), 1,
                ifelse(over == 2 & is.na(under) & is.na(average) & is.na(obes), 2,
                ifelse(obes == 3 & is.na(under) & is.na(over) & is.na(average), 3, NA)))))

HS12C$bmic <- as.factor(HS12C$bmic)

# exclude underweight
c12 <- subset(HS12C, !bmic == 0 | is.na(bmic))


## BMI
c12$bmic <- factor(c12$bmic, labels=c("Average","Overweight","Obese"))

prop_frame(c12, "bmic")
```

```r

## Sex
c12$gender <- factor(c12$gender,labels=c("Male","Female"))

prop_frame(c12, "gender")


## Age
summary(c12$age)


## Ethnicity
# recode anyone ~ european
c12 = c12 %>%
  mutate(euro = ifelse(C4_03_01 == 1, 1, 0))

# recode anyone ~ maori
c12 = c12 %>%
  mutate(maori = ifelse(C4_03_02 == 1, 1, 0))

# recode anyone ~ pacific
c12 = c12 %>%
  mutate(pacific = ifelse(C4_03_03 == 1, 1,
                   ifelse(C4_03_04 == 1, 1,
                   ifelse(C4_03_05 == 1, 1,
                   ifelse(C4_03_06 == 1, 1, 0)))))

# recode anyone ~ asian
# starting from 2011, other asian will be coded as other ethnicity because there is no way to identify other asian
c12 = c12 %>%
  mutate(asian = ifelse(C4_03_07 == 1, 1,
                 ifelse(C4_03_08 == 1, 1, 0)))

# recode anyone ~ other
c12 = c12 %>%
  mutate(other = ifelse(C4_03_77 == 1, 1, 0))


c12 = c12 %>%
```

```r
mutate(eth_count = ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "Maori Only",
                   ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Pacific Only",
                   ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Asian Only",
                   ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 0, "European Only",
                   ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other",
                   ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+ Ethnicities (M)",
                   ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "2+ Ethnicities (M)",
                   ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Other",
                   ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Other",
                   ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other",
                   ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                   ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                   ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other",
                   ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other",
                   ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other",
                   ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+ Ethnicities (M)",
                   ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                   ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                   ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other",
                   ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other",
```

```
                    ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other",
                    ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                    ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                    ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+ Ethnicities (M)",
                    ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other",
                    ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                    ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                    ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+ Ethnicities (M)",
                    ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other",
                    ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+ Ethnicities (M)",
                    ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+ Ethnicities (M)", NA)))))))))))))))))))))))))))))))

c12$eth_count <- as.factor(c12$eth_count)

prop_frame(c12, "eth_count")


## Household Income
# merge adult data to child data set
ac12 <- left_join(select(c12,HHID,C4_17), select(a12, HHID, A5_24, A5_14, A5_15), by="HHID")

ac12$ainc <- as.numeric(as.character(ac12$A5_24))
ac12$cinc <- as.numeric(as.character(ac12$C4_17))

ac12 <- mutate(ac12, inc= ifelse(is.na(ainc), cinc,
                             ifelse(is.na(cinc), ainc, cinc)))

ac12$hhinc <- cut(ac12$inc, breaks=c(0,5,10,13,16))
```

```
ac12$hhinc <- factor(ac12$hhinc, labels=c("<=$15,000","$15,001-$40,000","$40,001-$70,000",">$70,000"))

# copy it back to original data set
c12$hhinc <- ac12$hhinc

prop_frame(c12, "hhinc")


## Education
ac12$secondary <- as.numeric(as.character(ac12$A5_14))
ac12$secondary <- recode(ac12$secondary, "1" = "1", .default = "2") # 1= no degree, 2= secondary school grads
ac12$secondary <- as.factor(ac12$secondary)

ac12$tertiary <- as.numeric(as.character(ac12$A5_15))
ac12$tertiary <- recode(ac12$tertiary, "0" = "0", "1" = "2", "2" = "2",
                        "3" = "2", "4" = "2", .default = "3") #0 = no tertiary, 3= tertiary
ac12$tertiary <- as.factor(ac12$tertiary)

# prioritise highest education for duplicates
ac12 <- mutate(ac12, edu = ifelse(secondary == "1" & tertiary == "0", 1,
                           ifelse(secondary == "2" & tertiary == "0", 2,
                           ifelse(tertiary == "2", 2,
                           ifelse(tertiary == "3", 3, 1)))))

ac12$edu <- as.factor(ac12$edu)

ac12$edu <- factor(ac12$edu, labels=c("No Qualification","Secondary","Tertiary"))

# move them to the original data set
c12$edu <- ac12$edu

prop_frame(c12, "edu")

# remove ac12 from the memory
rm(ac12)


## Deprivation Quintile
```

```r
names(c12)[names(c12) == "nzdep06_quintile"] <- "dep"

prop_frame(c12, "dep")


## Fruit and Vegetable
summary(c12$kidfruit_2serves) # 5 K
summary(c12$kidveges_new) # 2K, 1R
c12$kidfruit_2serves[c12$kidfruit_2serves == "K"] <- NA
c12$kidfruit_2serves[c12$kidfruit_2serves == "X"] <- NA
c12$kidveges_new[c12$kidveges_new == "K"] <- NA
c12$kidveges_new[c12$kidveges_new == "R"] <- NA
c12$kidveges_new[c12$kidveges_new == "X"] <- NA

c12$fruit <- factor(c12$kidfruit_2serves, labels=c("No","Yes"))
c12$veges <- factor(c12$kidveges_new, labels=c("No","Yes"))

prop_frame(c12, "fruit")
prop_frame(c12, "veges")


## Soft Drink and Fast Food
summary(c12$C3_09) # 1 NA, 19 K, 1 R
c12$softd <-as.numeric(as.character(c12$C3_09))

c12$softd <- cut(c12$softd, c(0,1,2,4,Inf),labels=c("0/week",
                "1/week","2-3/week","4+/week"), right=FALSE)
summary(c12$softd)


summary(c12$C3_10) # 1 NA, 8 K, 1 R
c12$fastf <- as.numeric(as.character(c12$C3_10))

c12$fastf <- cut(c12$fastf, c(0,1,2,4,Inf), labels=c("0/week","1/week",
                 "2-3/week","4+/week"), right=FALSE)
summary(c12$fastf)

prop_frame(c12, "softd")
prop_frame(c12, "fastf")

```

```
#### NZHS 2013/14 - adult ####

summary(HS13A$bmi) # 261 R, 260 U, 194 P

HS13A$gender <- as.factor(HS13A$gender)

HS13A$bmiscale <- as.numeric(as.character(HS13A$bmi))

# Recode BMI category according to IOTF cut-offs
HS13A <- HS13A %>%
  mutate(bmic = ifelse(gender == 0 & age == 15 & bmiscale <16.98, 0,
                ifelse(gender == 0 & age == 16 & bmiscale <17.53, 0,
                ifelse(gender == 0 & age == 17 & bmiscale <18.04, 0,
                ifelse(gender == 1 & age == 15 & bmiscale <17.43, 0,
                ifelse(gender == 1 & age == 16 & bmiscale <17.9, 0,
                ifelse(gender == 1 & age == 17 & bmiscale <18.24, 0,
                ifelse(age >= 18 & bmiscale <18.5, 0,
                ifelse(gender == 0 & age == 15 & bmiscale >= 16.98 & bmiscale <23.28, 1,
                ifelse(gender == 0 & age == 16 & bmiscale >= 17.53 & bmiscale <23.89, 1,
                ifelse(gender == 0 & age == 17 & bmiscale >= 18.04 & bmiscale <24.46, 1,
                ifelse(gender == 1 & age == 15 & bmiscale >= 17.43 & bmiscale <23.89, 1,
                ifelse(gender == 1 & age == 16 & bmiscale >= 17.90 & bmiscale <24.34, 1,
                ifelse(gender == 1 & age == 17 & bmiscale >= 18.24 & bmiscale <24.70, 1,
                ifelse(age >= 18 & bmiscale >= 18.5 & bmiscale <25, 1,
                ifelse(gender == 0 & age == 15 & bmiscale >= 23.28 & bmiscale <28.32, 2,
                ifelse(gender == 0 & age == 16 & bmiscale >= 23.89 & bmiscale <28.89, 2,
                ifelse(gender == 0 & age == 17 & bmiscale >= 24.46 & bmiscale <29.43, 2,
                ifelse(gender == 1 & age == 15 & bmiscale >= 23.89 & bmiscale <29.01, 2,
                ifelse(gender == 1 & age == 16 & bmiscale >= 24.34 & bmiscale <29.40, 2,
                ifelse(gender == 1 & age == 17 & bmiscale >= 24.70 & bmiscale <29.70, 2,
                ifelse(age >= 18 & bmiscale >= 25 & bmiscale <30, 2,
                ifelse(gender == 0 & age == 15 & bmiscale >=28.32, 3,
                ifelse(gender == 0 & age == 16 & bmiscale >=28.89, 3,
                ifelse(gender == 0 & age == 17 & bmiscale >=29.43, 3,
                ifelse(gender == 1 & age == 15 & bmiscale >=29.01, 3,
                ifelse(gender == 1 & age == 16 & bmiscale >=29.40, 3,
                ifelse(gender == 1 & age == 17 & bmiscale >=29.70, 3,
                ifelse(age >= 18 & bmiscale >= 30, 3 ,NA))))))))))))))))))))))))))))
```

```
HS13A$bmic <- as.factor(HS13A$bmic)

# exclude Underweight
a13 <- subset(HS13A, !bmic == 0 | is.na(bmic))


## Migration Status
a13$native <- recode(a13$A5_06, "2004"="Migrant", "2005"="Migrant", "2006"="Migrant", "2007"="Migrant",
                                "2008"="Migrant", "2009"="Migrant", "2010"="Migrant", "2011"="Migrant",
                                "2012"="Migrant", "2013"="Migrant", "2014"="Migrant", ".K"= "S", ".R"= "S",
                                .default="Native")

a13$native[a13$native == "S"] <- NA  #recode S into NA
a13$native <- droplevels(a13$native)

prop_frame(a13, "native")


## BMI
a13$bmic <- factor(a13$bmic, labels=c("Average","Overweight","Obese"))

prop_frame(a13, "bmic")


### Sex
a13$gender <- factor(a13$gender, labels=c("Male","Female"))

prop_frame(a13, "gender")


### Age
summary(a13$age)


## Ethnicity
# recode anyone ~ european
```

```
a13 = a13 %>%
  mutate(euro = ifelse(A5_03_01 == 1, 1, 0))

# recode anyone ~ maori
a13 = a13 %>%
  mutate(maori = ifelse(A5_03_02 == 1, 1, 0))

# recode anyone ~ pacific
a13 = a13 %>%
  mutate(pacific = ifelse(A5_03_03 == 1, 1,
                   ifelse(A5_03_04 == 1, 1,
                   ifelse(A5_03_05 == 1, 1,
                   ifelse(A5_03_06 == 1, 1, 0)))))

# recode anyone ~ asian (Indian and Chinese)
a13 = a13 %>%
  mutate(asian = ifelse(A5_03_07 == 1, 1,
                 ifelse(A5_03_08 == 1, 1, 0)))

# recode anyone ~ other
a13 = a13 %>%
  mutate(other = ifelse(A5_03_77 == 1, 1, 0))


a13 = a13 %>%
  mutate(eth_count = ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "Maori
      Only",
                     ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "
                        Pacific Only",
                     ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Asian
                        Only",
                     ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 0, "
                        European Only",
                     ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other"
                        ,
                     ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+
                        Ethnicities (M)",
                     ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "2+
                        Ethnicities (M)",
```

```
                ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Other",
                ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Other",
                ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other",
                ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other",
                ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other",
                ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other",
                ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+ Ethnicities (M)",
                ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other",
                ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other",
                ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other",
                ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+ Ethnicities (M)",
                ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other",
                ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+ Ethnicities (M)",
```

```
                        ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+
                            Ethnicities (M)",
                        ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other"
                            ,
                        ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
                            Ethnicities (M)",
                        ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
                            Ethnicities (M)", NA)))))))))))))))))))))))))))))))

a13$eth_count <- as.factor(a13$eth_count)

prop_frame(a13, "eth_count")


## Household Income
summary(a13$A5_24)
a13$hhinc <- ordered(a13$A5_24, levels=c("4","5","6","7","8","9","10","11","12",
                                          "13","14","15","16"))

a13$hhinc <- as.numeric(as.character(a13$hhinc))
a13$hhinc <- cut(a13$hhinc, breaks=c(0,5,10,13,16))
a13$hhinc <- factor(a13$hhinc, labels=c("<=$15,000","$15,001-$40,000","$40,001-$70,000",">$70,000"))

prop_frame(a13, "hhinc")


## Education
a13$secondary <- as.numeric(as.character(a13$A5_14))
a13$secondary <- recode(a13$secondary, "1" = "1", .default = "2") # 1= no degree, 2= secondary
    school grads
a13$secondary <- as.factor(a13$secondary)

a13$tertiary <- as.numeric(as.character(a13$A5_15))
a13$tertiary <- recode(a13$tertiary, "0" = "0","1"="2","2"="2",
                       "3"="2", "4"="2", .default = "3") #0 = no tertiary, 3= tertiary
a13$tertiary <- as.factor(a13$tertiary)

a13 <- mutate(a13, edu = ifelse(secondary =="1" & tertiary =="0", 1,
                          ifelse(secondary == "2" & tertiary == "0", 2,
                          ifelse(tertiary == "2", 2,
```

```
                          ifelse(tertiary == "3", 3, 1)))))

a13$edu <- as.factor(a13$edu)
a13$edu <- factor(a13$edu, labels=c("No Qualification","Secondary","Tertiary"))

prop_frame(a13, "edu")


## Deprivation Quintile
names(a13)[names(a13) == "nzdep13_quintile"] <- "dep"

prop_frame(a13, "dep")


## Fruit and Vegetable
summary(a13$fruit)
summary(a13$veges)
a13$fruit[a13$fruit =="K"] <- NA
a13$fruit[a13$fruit =="R"] <- NA
a13$veges[a13$veges =="K"] <- NA
a13$veges[a13$veges =="R"] <- NA

a13$fruit <- factor(a13$fruit, labels=c("No","Yes"))
a13$veges <- factor(a13$veges, labels=c("No","Yes"))

prop_frame(a13, "fruit")
prop_frame(a13, "veges")


## Smoking Status
a13 <- a13 %>%
  mutate(smoke = ifelse(current_smoker == 1 & ex_smoker == 0, 3,
                ifelse(current_smoker == 0 & ex_smoker == 1, 2,
                 ifelse(current_smoker == "K", NA,
                 ifelse(current_smoker == "R", NA,
                 ifelse(ex_smoker =="K", NA,
                 ifelse(ex_smoker =="R", NA, 1)))))))

a13$smoke <- as.factor(a13$smoke)
a13$smoke <- factor(a13$smoke, labels=c("Non Smoker","Ex Smoker","Current Smoker"))
```

```
prop_frame(a13, "smoke")


## Alcohol Problem
a13$haz_drinker_all[a13$haz_drinker_all == "K"] <- NA
a13$haz_drinker_all[a13$haz_drinker_all == "R"] <- NA

a13$haz_drinker_all <- factor(a13$haz_drinker_all, labels=c("No Alcohol Problem",
                                                            "Alcohol Problem"))

prop_frame(a13, "haz_drinker_all")


## Physical Activity
a13$active[a13$active == "K"] <- NA
a13$active[a13$active == "R"] <- NA

a13$active <- factor(a13$active, labels=c("Not Active","Active"))

prop_frame(a13, "active")


## Sedentary Lifestyle
a13$sedentary[a13$sedentary == "K"] <- NA
a13$sedentary[a13$sedentary == "R"] <- NA

a13$sedentary <- factor(a13$sedentary, labels=c("Not Sedentary","Sedentary"))

prop_frame(a13, "sedentary")


# Difficulty Climbing Several Steps of Stairs
summary(a13$A4_03)
a13$A4_03[a13$A4_03 == "."] <- NA
a13$A4_03[a13$A4_03 == ".K"] <- NA
a13$A4_03[a13$A4_03 == ".R"] <- NA

a13$stair <- factor(a13$A4_03, labels=c("A Lot Difficult","A Little Difficult",
                                        "No Difficulty"))
```

```
prop_frame(a13, "stair")



#### NZHS 2013/14 — child ####

summary(HS13C$bmi) # 981 U, 140 R
HS13C$bmiscale <- as.numeric(as.character(HS13C$bmi))

HS13C$gender <- as.factor(HS13C$gender)

# Recode BMI category based on IOTF cut-offs
HS13C <- HS13C %>%
  mutate(under =
  ifelse(gender == 0 & age == 2 & bmiscale <15.24, 0,
  ifelse(gender == 0 & age == 3 & bmiscale <14.83, 0,
  ifelse(gender == 0 & age == 4 & bmiscale <14.51, 0,
  ifelse(gender == 0 & age == 5 & bmiscale <14.26, 0,
  ifelse(gender == 0 & age == 6 & bmiscale <14.06, 0,
  ifelse(gender == 0 & age == 7 & bmiscale <14.00, 0,
  ifelse(gender == 0 & age == 8 & bmiscale <14.13, 0,
  ifelse(gender == 0 & age == 9 & bmiscale <14.36, 0,
  ifelse(gender == 0 & age == 10 & bmiscale <14.63, 0,
  ifelse(gender == 0 & age == 11 & bmiscale <14.96, 0,
  ifelse(gender == 0 & age == 12 & bmiscale <15.36, 0,
  ifelse(gender == 0 & age == 13 & bmiscale <15.84, 0,
  ifelse(gender == 0 & age == 14 & bmiscale <16.39, 0,
  ifelse(gender == 0 & age == 15 & bmiscale <16.98, 0,
  ifelse(gender == 0 & age == 16 & bmiscale <17.53, 0,
  ifelse(gender == 0 & age == 17 & bmiscale <18.04, 0,
  ifelse(gender == 1 & age == 2 & bmiscale <14.96, 0,
  ifelse(gender == 1 & age == 3 & bmiscale <14.60, 0,
  ifelse(gender == 1 & age == 4 & bmiscale <14.30, 0,
  ifelse(gender == 1 & age == 5 & bmiscale <14.04, 0,
  ifelse(gender == 1 & age == 6 & bmiscale <13.85, 0,
  ifelse(gender == 1 & age == 7 & bmiscale <13.83, 0,
  ifelse(gender == 1 & age == 8 & bmiscale <14.00, 0,
  ifelse(gender == 1 & age == 9 & bmiscale <14.26, 0,
  ifelse(gender == 1 & age == 10 & bmiscale <14.58, 0,
```

```
  ifelse(gender == 1 & age == 11 & bmiscale <15.03, 0,
  ifelse(gender == 1 & age == 12 & bmiscale <15.59, 0,
  ifelse(gender == 1 & age == 13 & bmiscale <16.23, 0,
  ifelse(gender == 1 & age == 14 & bmiscale <16.86, 0,
  ifelse(gender == 1 & age == 15 & bmiscale <17.43, 0,
  ifelse(gender == 1 & age == 16 & bmiscale <17.90, 0,
  ifelse(gender == 1 & age == 17 & bmiscale <18.24, 0,
  ifelse(age >= 18 & bmiscale <18.5, 0, NA))))))))))))))))))))))))))))))))

HS13C <- HS13C %>%
  mutate(average =
  ifelse(gender == 0 & age == 2 & bmiscale >=15.24 & bmiscale <18.36, 1,
  ifelse(gender == 0 & age == 3 & bmiscale >=14.83 & bmiscale <17.85, 1,
  ifelse(gender == 0 & age == 4 & bmiscale >=14.51 & bmiscale <17.52, 1,
  ifelse(gender == 0 & age == 5 & bmiscale >=14.26 & bmiscale <17.39, 1,
  ifelse(gender == 0 & age == 6 & bmiscale >=14.06 & bmiscale <17.52, 1,
  ifelse(gender == 0 & age == 7 & bmiscale >=14.00 & bmiscale <17.88, 1,
  ifelse(gender == 0 & age == 8 & bmiscale >=14.13 & bmiscale <18.41, 1,
  ifelse(gender == 0 & age == 9 & bmiscale >=14.36 & bmiscale <19.07, 1,
  ifelse(gender == 0 & age == 10 & bmiscale >=14.63 & bmiscale <19.80, 1,
  ifelse(gender == 0 & age == 11 & bmiscale >=14.96 & bmiscale <20.51, 1,
  ifelse(gender == 0 & age == 12 & bmiscale >=15.36 & bmiscale <21.20, 1,
  ifelse(gender == 0 & age == 13 & bmiscale >=15.84 & bmiscale <21.89, 1,
  ifelse(gender == 0 & age == 14 & bmiscale >=16.39 & bmiscale <22.60, 1,
  ifelse(gender == 0 & age == 15 & bmiscale >=16.98 & bmiscale <23.28, 1,
  ifelse(gender == 0 & age == 16 & bmiscale >=17.53 & bmiscale <23.89, 1,
  ifelse(gender == 0 & age == 17 & bmiscale >=18.04 & bmiscale <24.46, 1,
  ifelse(gender == 1 & age == 2 & bmiscale >=14.96 & bmiscale <18.09, 1,
  ifelse(gender == 1 & age == 3 & bmiscale >=14.60 & bmiscale <17.64, 1,
  ifelse(gender == 1 & age == 4 & bmiscale >=14.30 & bmiscale <17.35, 1,
  ifelse(gender == 1 & age == 5 & bmiscale >=14.04 & bmiscale <17.23, 1,
  ifelse(gender == 1 & age == 6 & bmiscale >=13.85 & bmiscale <17.33, 1,
  ifelse(gender == 1 & age == 7 & bmiscale >=13.83 & bmiscale <17.69, 1,
  ifelse(gender == 1 & age == 8 & bmiscale >=14.00 & bmiscale <18.28, 1,
  ifelse(gender == 1 & age == 9 & bmiscale >=14.26 & bmiscale <18.99, 1,
  ifelse(gender == 1 & age == 10 & bmiscale >=14.58 & bmiscale <19.78, 1,
  ifelse(gender == 1 & age == 11 & bmiscale >=15.03 & bmiscale <20.66, 1,
  ifelse(gender == 1 & age == 12 & bmiscale >=15.59 & bmiscale <21.59, 1,
  ifelse(gender == 1 & age == 13 & bmiscale >=16.23 & bmiscale <22.49, 1,
  ifelse(gender == 1 & age == 14 & bmiscale >=16.86 & bmiscale <23.27, 1,
```

```
  ifelse(gender == 1 & age == 15 & bmiscale >=17.43 & bmiscale <23.89, 1,
  ifelse(gender == 1 & age == 16 & bmiscale >=17.90 & bmiscale <24.34, 1,
  ifelse(gender == 1 & age == 17 & bmiscale >=18.24 & bmiscale <24.70, 1,
  ifelse(age >= 18 & bmiscale >=18.5 & bmiscale <25, 1, NA))))))))))))))))))))))))))))))))))

HS13C <- HS13C %>%
  mutate(over=
  ifelse(gender == 0 & age == 2 & bmiscale >=18.36 & bmiscale <19.99, 2,
  ifelse(gender == 0 & age == 3 & bmiscale >=17.85 & bmiscale <19.50, 2,
  ifelse(gender == 0 & age == 4 & bmiscale >=17.52 & bmiscale <19.23, 2,
  ifelse(gender == 0 & age == 5 & bmiscale >=17.39 & bmiscale <19.27, 2,
  ifelse(gender == 0 & age == 6 & bmiscale >=17.52 & bmiscale <19.76, 2,
  ifelse(gender == 0 & age == 7 & bmiscale >=17.88 & bmiscale <20.59, 2,
  ifelse(gender == 0 & age == 8 & bmiscale >=18.41 & bmiscale <21.56, 2,
  ifelse(gender == 0 & age == 9 & bmiscale >=19.07 & bmiscale <22.71, 2,
  ifelse(gender == 0 & age == 10 & bmiscale >=19.80 & bmiscale <23.96, 2,
  ifelse(gender == 0 & age == 11 & bmiscale >=20.51 & bmiscale <25.07, 2,
  ifelse(gender == 0 & age == 12 & bmiscale >=21.20 & bmiscale <26.02, 2,
  ifelse(gender == 0 & age == 13 & bmiscale >=21.89 & bmiscale <26.87, 2,
  ifelse(gender == 0 & age == 14 & bmiscale >=22.60 & bmiscale <27.64, 2,
  ifelse(gender == 0 & age == 15 & bmiscale >=23.28 & bmiscale <28.32, 2,
  ifelse(gender == 0 & age == 16 & bmiscale >=23.89 & bmiscale <28.89, 2,
  ifelse(gender == 0 & age == 17 & bmiscale >=24.46 & bmiscale <29.43, 2,
  ifelse(gender == 1 & age == 2 & bmiscale >=18.09 & bmiscale <19.81, 2,
  ifelse(gender == 1 & age == 3 & bmiscale >=17.64 & bmiscale <19.38, 2,
  ifelse(gender == 1 & age == 4 & bmiscale >=17.35 & bmiscale <19.16, 2,
  ifelse(gender == 1 & age == 5 & bmiscale >=17.23 & bmiscale <19.20, 2,
  ifelse(gender == 1 & age == 6 & bmiscale >=17.33 & bmiscale <19.61, 2,
  ifelse(gender == 1 & age == 7 & bmiscale >=17.69 & bmiscale <20.39, 2,
  ifelse(gender == 1 & age == 8 & bmiscale >=18.28 & bmiscale <21.44, 2,
  ifelse(gender == 1 & age == 9 & bmiscale >=18.99 & bmiscale <22.66, 2,
  ifelse(gender == 1 & age == 10 & bmiscale >=19.78 & bmiscale <23.97, 2,
  ifelse(gender == 1 & age == 11 & bmiscale >=20.66 & bmiscale <25.25, 2,
  ifelse(gender == 1 & age == 12 & bmiscale >=21.59 & bmiscale <26.47, 2,
  ifelse(gender == 1 & age == 13 & bmiscale >=22.49 & bmiscale <27.57, 2,
  ifelse(gender == 1 & age == 14 & bmiscale >=23.27 & bmiscale <28.42, 2,
  ifelse(gender == 1 & age == 15 & bmiscale >=23.89 & bmiscale <29.01, 2,
  ifelse(gender == 1 & age == 16 & bmiscale >=24.34 & bmiscale <29.40, 2,
  ifelse(gender == 1 & age == 17 & bmiscale >=24.70 & bmiscale <29.70, 2,
  ifelse(age >= 18 & bmiscale >=25 & bmiscale <30, 2, NA))))))))))))))))))))))))))))))))))
```

```
HS13C <- HS13C %>%
  mutate(obes=
  ifelse(gender == 0 & age == 2 & bmiscale >=19.99, 3,
  ifelse(gender == 0 & age == 3 & bmiscale >=19.50, 3,
  ifelse(gender == 0 & age == 4 & bmiscale >=19.23, 3,
  ifelse(gender == 0 & age == 5 & bmiscale >=19.27, 3,
  ifelse(gender == 0 & age == 6 & bmiscale >=19.76, 3,
  ifelse(gender == 0 & age == 7 & bmiscale >=20.59, 3,
  ifelse(gender == 0 & age == 8 & bmiscale >=21.56, 3,
  ifelse(gender == 0 & age == 9 & bmiscale >=22.71, 3,
  ifelse(gender == 0 & age == 10 & bmiscale >=23.96, 3,
  ifelse(gender == 0 & age == 11 & bmiscale >=25.07, 3,
  ifelse(gender == 0 & age == 12 & bmiscale >=26.02, 3,
  ifelse(gender == 0 & age == 13 & bmiscale >=26.87, 3,
  ifelse(gender == 0 & age == 14 & bmiscale >=27.64, 3,
  ifelse(gender == 0 & age == 15 & bmiscale >=28.32, 3,
  ifelse(gender == 0 & age == 16 & bmiscale >=28.89, 3,
  ifelse(gender == 0 & age == 17 & bmiscale >=29.43, 3,
  ifelse(gender == 1 & age == 2 & bmiscale >=19.81, 3,
  ifelse(gender == 1 & age == 3 & bmiscale >=19.38, 3,
  ifelse(gender == 1 & age == 4 & bmiscale >=19.16, 3,
  ifelse(gender == 1 & age == 5 & bmiscale >=19.20, 3,
  ifelse(gender == 1 & age == 6 & bmiscale >=19.61, 3,
  ifelse(gender == 1 & age == 7 & bmiscale >=20.39, 3,
  ifelse(gender == 1 & age == 8 & bmiscale >=21.44, 3,
  ifelse(gender == 1 & age == 9 & bmiscale >=22.66, 3,
  ifelse(gender == 1 & age == 10 & bmiscale >=23.97, 3,
  ifelse(gender == 1 & age == 11 & bmiscale >=25.25, 3,
  ifelse(gender == 1 & age == 12 & bmiscale >=26.47, 3,
  ifelse(gender == 1 & age == 13 & bmiscale >=27.57, 3,
  ifelse(gender == 1 & age == 14 & bmiscale >=28.42, 3,
  ifelse(gender == 1 & age == 15 & bmiscale >=29.01, 3,
  ifelse(gender == 1 & age == 16 & bmiscale >=29.40, 3,
  ifelse(gender == 1 & age == 17 & bmiscale >=29.70, 3,
  ifelse(age >= 18 & bmiscale >=30, 3, NA))))))))))))))))))))))))))))))))))


# 0: underweight, 1: average, 2: overweight, 3: obese
HS13C <- HS13C %>%
```

```
  mutate(bmic = ifelse(under == 0 & is.na(average) & is.na(over) & is.na(obes), 0,
                ifelse(average == 1 & is.na(under) & is.na(over) & is.na(obes), 1,
                ifelse(over == 2 & is.na(under) & is.na(average) & is.na(obes), 2,
                ifelse(obes == 3 & is.na(under) & is.na(over) & is.na(average), 3, NA)))))

HS13C$bmic <- as.factor(HS13C$bmic)

# Exclude underweight
c13 <- subset(HS13C, !bmic == 0 | is.na(bmic))


## BMI
c13$bmic <- factor(c13$bmic, labels=c("Average","Overweight","Obese"))

prop_frame(c13, "bmic")


## Sex
c13$gender <- factor(c13$gender, labels=c("Male","Female"))

prop_frame(c13, "gender")


## Age
summary(c13$age)


## Ethnicity
# recode anyone ~ european
c13 = c13 %>%
  mutate(euro = ifelse(C4_03_01 == 1, 1, 0))

# recode anyone ~ maori
c13 = c13 %>%
  mutate(maori = ifelse(C4_03_02 == 1, 1, 0))

# recode anyone ~ pacific
c13 = c13 %>%
  mutate(pacific = ifelse(C4_03_03 == 1, 1,
                   ifelse(C4_03_04 == 1, 1,
```

```
                     ifelse(C4_03_05 == 1, 1,
                     ifelse(C4_03_06 == 1, 1, 0))))))

# recode anyone ~ asian
c13 = c13 %>%
  mutate(asian = ifelse(C4_03_07 == 1, 1,
                    ifelse(C4_03_08 == 1, 1, 0)))

# recode anyone ~ other
c13 = c13 %>%
  mutate(other = ifelse(C4_03_77 == 1, 1, 0))


c13 = c13 %>%
  mutate(eth_count = ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "Maori Only",
                       ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Pacific Only",
                       ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Asian Only",
                       ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 0, "European Only",
                       ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other",
                       ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+ Ethnicities (M)",
                       ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "2+ Ethnicities (M)",
                       ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Other",
                       ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Other",
                       ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other",
                       ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                       ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                       ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other",
```

```
                        ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other",
                        ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other",
                        ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+ Ethnicities (M)",
                        ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                        ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                        ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other",
                        ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other",
                        ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other",
                        ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                        ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                        ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+ Ethnicities (M)",
                        ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other",
                        ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                        ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                        ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+ Ethnicities (M)",
                        ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other",
                        ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+ Ethnicities (M)",
                        ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+ Ethnicities (M)", NA))))))))))))))))))))))))))))))))

c13$eth_count <- as.factor(c13$eth_count)

prop_frame(c13, "eth_count")
```

```r


## Household Income
# combined adult to child data set to match income and education data
ac13 <- left_join(select(c13, HHID, C4_17, C4_20, C4_21, CQ1, CQ2),
                  select(a13, HHID, A5_24, A5_14, A5_15), by="HHID")

summary(ac13$A5_24)
ac13$hhinc1 <- ordered(ac13$A5_24, levels=c("4","5","6","7","8","9","10","11","12",
                                            "13","14","15","16"))
summary(ac13$C4_17)
ac13$hhinc2 <- ordered(ac13$C4_17, levels=c("4","5","6","7","8","9","10","11","12",
                                            "13","14","15","16"))

ac13$hhinc1 <-as.numeric(as.character(ac13$hhinc1))
ac13$hhinc2 <-as.numeric(as.character(ac13$hhinc2))

ac13$hhinc1 <- cut(ac13$hhinc1, breaks=c(0,5,10,13,16))
ac13$hhinc2 <- cut(ac13$hhinc2, breaks=c(0,5,10,13,16))

ac13$hhinc1 <- factor(ac13$hhinc1, labels=c(1,2,3,4))
ac13$hhinc2 <- factor(ac13$hhinc2, labels=c(1,2,3,4))

ac13 <- mutate(ac13, hhinc= ifelse(hhinc1 == 1 & CQ1 == 1, 1,
                            ifelse(hhinc1 == 2 & CQ1 == 1, 2,
                            ifelse(hhinc1 == 3 & CQ1 == 1, 3,
                            ifelse(hhinc1 == 4 & CQ1 == 1, 4,
                            ifelse(hhinc2 == 1 & CQ1 == 2, 1,
                            ifelse(hhinc2 == 2 & CQ1 == 2, 2,
                            ifelse(hhinc2 == 3 & CQ1 == 2, 3,
                            ifelse(hhinc2 == 4 & CQ1 == 2, 4, NA)))))))))

ac13$hhinc <- as.factor(ac13$hhinc)

# recode back into child data set
ac13$hhinc <- factor(ac13$hhinc, labels=c("<=$15,000",
             "$15,001-$40,000","$40,001-$70,000",">$70,000"))
c13$hhinc <- ac13$hhinc

prop_frame(c13, "hhinc")
```

```
## Education
ac13$cedu1 <- as.numeric(as.character(ac13$C4_20))
ac13$aedu1 <- as.numeric(as.character(ac13$A5_14))

ac13 <- mutate(ac13, edu1 = ifelse(is.na(cedu1), aedu1,
                            ifelse(is.na(aedu1), cedu1,
                            ifelse(cedu1 >= aedu1, cedu1,
                            ifelse(aedu1 >= cedu1, aedu1, NA)))))

ac13$edu1 <- as.factor(ac13$edu1)
summary(ac13$edu1)

# coding discrepancies, decided to use the highest qualification in the household
ac13$secondary <- as.numeric(as.character(ac13$edu1))
ac13$secondary <- recode(ac13$secondary, "1" = "1", .default = "2") # 1= no degree, 2= secondary
    school grads
ac13$secondary <- as.factor(ac13$secondary)

summary(ac13$C4_21) # used different coding than the adult set
summary(ac13$A5_15)

ac13$C4_21[ac13$C4_21 == "77"] <- NA
ac13$A5_15[ac13$A5_15 == "77"] <- NA

ac13$cedu2 <- as.numeric(as.character(ac13$C4_21))-1 # to match the coding in the adult set
ac13$aedu2 <- as.numeric(as.character(ac13$A5_15))

ac13 <- mutate(ac13, edu2 = ifelse(is.na(cedu2), aedu2,
                            ifelse(is.na(aedu2), cedu2,
                            ifelse(aedu2 >= cedu2, aedu2,
                            ifelse(cedu2 >= aedu2, cedu2, NA)))))

ac13$edu2 <- as.factor(ac13$edu2)
summary(ac13$edu2)


ac13$tertiary <- as.numeric(as.character(ac13$edu2))
ac13$tertiary <- recode(ac13$tertiary, "0" = "0","1"="2","2"="2",
```

```
                            "3"="2", "4"="2", .default = "3") #0 = no tertiary, 3= tertiary
ac13$tertiary <- as.factor(ac13$tertiary)

# recode into one variable, prioritise highest education
ac13 <- mutate(ac13, edu = ifelse(secondary =="1" & tertiary =="0", 1,
                              ifelse(secondary == "2" & tertiary == "0", 2,
                              ifelse(tertiary == "2", 2,
                              ifelse(tertiary == "3", 3, 1)))))

ac13$edu <- factor(ac13$edu, labels=c("No Qualification","Secondary","Tertiary"))

# move them back into child data set
c13$edu <- ac13$edu

prop_frame(c13, "edu")

# remove ac13 from memory
rm(ac13)


## Deprivation Quintile
names(c13)[names(c13) == "nzdep13_quintile"] <- "dep"

prop_frame(c13, "dep")


## Fruit and Vegetable
c13$fruit <- c13$kidfruit_2serves
c13$veges <- c13$kidveges_new

c13$fruit[c13$fruit == "K"] <- NA
c13$fruit[c13$fruit == "X"] <- NA
c13$veges[c13$veges == "K"] <- NA
c13$veges[c13$veges == "R"] <- NA
c13$veges[c13$veges == "X"] <- NA

c13$fruit <- factor(c13$fruit, labels=c("No","Yes"))
c13$veges <- factor(c13$veges, labels=c("No","Yes"))

prop_frame(c13, "fruit")
```

```r
prop_frame(c13, "veges")


### Soft Drink and Fast Food
summary(c13$C3_09) # 2 NA, 16 K
c13$softd <- as.numeric(as.character(c13$C3_09)) # force non-numeric into NA
c13$softd <- cut(c13$softd, c(0,1,2,4,Inf), labels=c("0/week",
                 "1/week","2-3/week","4+/week"), right=FALSE)

summary(c13$C3_10)
c13$fastf <- as.numeric(as.character(c13$C3_10)) # force non-numeric into NA
c13$fastf <- cut(c13$fastf, c(0,1,2,4,Inf), labels=c("0/week",
                 "1/week","2-3/week","4+/week"), right=FALSE)

prop_frame(c13, "softd")
prop_frame(c13, "fastf")




#### NZHS 2014/15 - adult ####

summary(HS14A$bmi) # 527 R, 201 P

HS14A$gender <- as.factor(HS14A$gender)

HS14A$bmiscale <- as.numeric(as.character(HS14A$bmi))

# recode BMI category using IOTF cut-offs
HS14A <- HS14A %>%
  mutate(bmic = ifelse(gender == 0 & age == 15 & bmiscale <16.98, 0,
                ifelse(gender == 0 & age == 16 & bmiscale <17.53, 0,
                ifelse(gender == 0 & age == 17 & bmiscale <18.04, 0,
                ifelse(gender == 1 & age == 15 & bmiscale <17.43, 0,
                ifelse(gender == 1 & age == 16 & bmiscale <17.9, 0,
                ifelse(gender == 1 & age == 17 & bmiscale <18.24, 0,
                ifelse(age >= 18 & bmiscale <18.5, 0,
                ifelse(gender == 0 & age == 15 & bmiscale >= 16.98 & bmiscale <23.28, 1,
                ifelse(gender == 0 & age == 16 & bmiscale >= 17.53 & bmiscale <23.89, 1,
                ifelse(gender == 0 & age == 17 & bmiscale >= 18.04 & bmiscale <24.46, 1,
```

```
                    ifelse(gender == 1 & age == 15 & bmiscale >= 17.43 & bmiscale <23.89, 1,
                    ifelse(gender == 1 & age == 16 & bmiscale >= 17.90 & bmiscale <24.34, 1,
                    ifelse(gender == 1 & age == 17 & bmiscale >= 18.24 & bmiscale <24.70, 1,
                    ifelse(age >= 18 & bmiscale >= 18.5 & bmiscale <25, 1,
                    ifelse(gender == 0 & age == 15 & bmiscale >= 23.28 & bmiscale <28.32, 2,
                    ifelse(gender == 0 & age == 16 & bmiscale >= 23.89 & bmiscale <28.89, 2,
                    ifelse(gender == 0 & age == 17 & bmiscale >= 24.46 & bmiscale <29.43, 2,
                    ifelse(gender == 1 & age == 15 & bmiscale >= 23.89 & bmiscale <29.01, 2,
                    ifelse(gender == 1 & age == 16 & bmiscale >= 24.34 & bmiscale <29.40, 2,
                    ifelse(gender == 1 & age == 17 & bmiscale >= 24.70 & bmiscale <29.70, 2,
                    ifelse(age >= 18 & bmiscale >= 25 & bmiscale <30, 2,
                    ifelse(gender == 0 & age == 15 & bmiscale >=28.32, 3,
                    ifelse(gender == 0 & age == 16 & bmiscale >=28.89, 3,
                    ifelse(gender == 0 & age == 17 & bmiscale >=29.43, 3,
                    ifelse(gender == 1 & age == 15 & bmiscale >=29.01, 3,
                    ifelse(gender == 1 & age == 16 & bmiscale >=29.40, 3,
                    ifelse(gender == 1 & age == 17 & bmiscale >=29.70, 3,
                    ifelse(age >= 18 & bmiscale >= 30, 3 ,NA))))))))))))))))))))))))))))

HS14A$bmic <- as.factor(HS14A$bmic)
summary(HS14A$bmic)

# exclude underweight
a14 <- subset(HS14A, !bmic == 0 | is.na(bmic))


## Migration Status
summary(a14$A5_06)

a14$native <- as.numeric(as.character(a14$A5_06))
a14$native <- cut(a14$native, c(0,2005,Inf), right=FALSE)
summary(a14$native)

a14$native <- factor(a14$native, labels= c("Native","Migrant"))
a14$native[is.na(a14$native)] <- "Native" # convert all NA into native

prop_frame(a14, "native")


## BMI
```

```r
a14$bmic <- factor(a14$bmic, labels=c("Average","Overweight","Obese"))

prop_frame(a14, "bmic")


## Sex
a14$gender <- factor(a14$gender, labels=c("Male","Female"))

prop_frame(a14, "gender")


## Age
summary(a14$age)


## Ethnicity
# recode anyone ~ european
a14 = a14 %>%
  mutate(euro = ifelse(A5_03_01 == 1, 1, 0))

# recode anyone ~ maori
a14 = a14 %>%
  mutate(maori = ifelse(A5_03_02 == 1, 1, 0))

# recode anyone ~ pacific
a14 = a14 %>%
  mutate(pacific = ifelse(A5_03_03 == 1, 1,
                   ifelse(A5_03_04 == 1, 1,
                   ifelse(A5_03_05 == 1, 1,
                   ifelse(A5_03_06 == 1, 1, 0)))))

# recode anyone ~ asian (Indian and Chinese)
a14 = a14 %>%
  mutate(asian = ifelse(A5_03_07 == 1, 1,
                 ifelse(A5_03_08 == 1, 1, 0)))

# recode anyone ~ other
a14 = a14 %>%
  mutate(other = ifelse(A5_03_77 == 1, 1, 0))

```

```
a14 = a14 %>%
  mutate(eth_count = ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "Maori Only",
                     ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Pacific Only",
                     ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Asian Only",
                     ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 0, "European Only",
                     ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other",
                     ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+ Ethnicities (M)",
                     ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "2+ Ethnicities (M)",
                     ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Other",
                     ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Other",
                     ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other",
                     ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                     ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                     ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other",
                     ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other",
                     ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other",
                     ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+ Ethnicities (M)",
                     ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                     ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                     ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other",
```

```
                          ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other",
                          ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other",
                          ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                          ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                          ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+ Ethnicities (M)",
                          ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other",
                          ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                          ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                          ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+ Ethnicities (M)",
                          ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other",
                          ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+ Ethnicities (M)",
                          ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+ Ethnicities (M)", NA)))))))))))))))))))))))))))))))

a14$eth_count <- as.factor(a14$eth_count)

prop_frame(a14, "eth_count")


## Household Income
summary(a14$A5_24)
a14$hhinc <- ordered(a14$A5_24, levels=c("4","5","6","7","8","9","10","11","12",
                                          "13","14","15","16"))

a14$hhinc <- as.numeric(as.character(a14$hhinc)) # force non-numeric into NA
a14$hhinc <- cut(a14$hhinc, breaks=c(0,5,10,13,16))
a14$hhinc <- factor(a14$hhinc, labels=c("<=$15,000","$15,001-$40,000","$40,001-$70,000",">$70,000"))

prop_frame(a14, "hhinc")
```

```
## Education
a14$secondary <- as.numeric(as.character(a14$A5_14)) # force non-numeric into NA
a14$secondary <- recode(a14$secondary, "1" = "1", .default = "2") # 1= no degree, 2= secondary
    school grads
a14$secondary <- as.factor(a14$secondary)

a14$tertiary <- as.numeric(as.character(a14$A5_15)) # force non-numeric into NA
a14$tertiary <- recode(a14$tertiary, "0" = "0","1"="2","2"="2",
                        "3"="2", "4"="2", .default = "3") #0 = no tertiary, 3= tertiary
a14$tertiary <- as.factor(a14$tertiary)

a14 <- mutate(a14, edu = ifelse(secondary =="1" & tertiary =="0", 1,
                         ifelse(secondary == "2" & tertiary == "0", 2,
                         ifelse(tertiary == "2", 2,
                         ifelse(tertiary == "3", 3, 1)))))

a14$edu <- as.factor(a14$edu)

a14$edu <- factor(a14$edu, labels=c("No Qualification","Secondary","Tertiary"))

prop_frame(a14, "edu")


## Deprivation Quintile
names(a14)[names(a14) == "nzdep13_quintile"] <- "dep"

prop_frame(a14, "dep")


## Fruit and Vegetable
a14$fruit[a14$fruit == "K"] <- NA
a14$veges[a14$veges == "K"] <- NA
a14$fruit <- factor(a14$fruit, labels=c("No","Yes"))
a14$veges <- factor(a14$veges, labels=c("No","Yes"))

prop_frame(a14, "fruit")
prop_frame(a14, "veges")

```

```
## Alcohol Problem
a14$haz_drinker_all[a14$haz_drinker_all == "K"] <- NA
a14$haz_drinker_all[a14$haz_drinker_all == "R"] <- NA

a14$haz_drinker_all <- factor(a14$haz_drinker_all, labels=c("No Alcohol Problem",
                                                            "Alcohol Problem"))

prop_frame(a14, "haz_drinker_all")


## Smoking Status
a14 <- a14 %>%
  mutate(smoke = ifelse(current_smoker == 1 & ex_smoker == 0, 3,
                 ifelse(current_smoker == 0 & ex_smoker == 1, 2,
                  ifelse(current_smoker == "K", NA,
                  ifelse(current_smoker == "R", NA,
                  ifelse(ex_smoker =="K", NA,
                  ifelse(ex_smoker =="R", NA, 1)))))))

a14$smoke <- as.factor(a14$smoke)
a14$smoke <- factor(a14$smoke, labels=c("Non Smoker","Ex Smoker","Current Smoker"))
summary(a14$smoke)

prop_frame(a14, "smoke")


## Physical ACtivity
a14$active[a14$active =="K"] <- NA
a14$active[a14$active =="R"] <- NA
a14$active <- factor(a14$active, labels=c("Not Active","Active"))

prop_frame(a14, "active")


## Sedentary Lifestyle
a14$sedentary[a14$sedentary =="K"] <- NA
a14$sedentary[a14$sedentary =="R"] <- NA
a14$sedentary <- factor(a14$sedentary, labels=c("Not Sedentary","Sedentary"))
```

```r
prop_frame(a14, "sedentary")


## Urban/Rural Area
a14$UA <- recode(a14$UrbanArea2015Class, "M"="Urban", "S"="Urban", .default="Rural")

prop_frame(a14, "UA")


## Difficulty Climbing Several Steps of Stairs
summary(a14$A4_03)
a14$A4_03[a14$A4_03 ==".K"] <- NA
a14$A4_03[a14$A4_03 ==".R"] <- NA

a14$stair <- factor(a14$A4_03, labels=c("A Lot Difficult","A Little Difficult",
                                        "No Difficulty"))

prop_frame(a14, "stair")



#### NZHS 2014/15 - child ####

summary(HS14C$bmi) # 981 U, 276 R, 6 O

HS14C$bmiscale <- as.numeric(as.character(HS14C$bmi))

HS14C$gender <- as.factor(HS14C$gender)

# recode into BMI Category based on IOTF cut-offs
HS14C <- HS14C %>%
  mutate(under =
  ifelse(gender == 0 & age == 2 & bmiscale <15.24, 0,
  ifelse(gender == 0 & age == 3 & bmiscale <14.83, 0,
  ifelse(gender == 0 & age == 4 & bmiscale <14.51, 0,
  ifelse(gender == 0 & age == 5 & bmiscale <14.26, 0,
  ifelse(gender == 0 & age == 6 & bmiscale <14.06, 0,
  ifelse(gender == 0 & age == 7 & bmiscale <14.00, 0,
  ifelse(gender == 0 & age == 8 & bmiscale <14.13, 0,
  ifelse(gender == 0 & age == 9 & bmiscale <14.36, 0,
```

```
  ifelse(gender == 0 & age == 10 & bmiscale <14.63, 0,
  ifelse(gender == 0 & age == 11 & bmiscale <14.96, 0,
  ifelse(gender == 0 & age == 12 & bmiscale <15.36, 0,
  ifelse(gender == 0 & age == 13 & bmiscale <15.84, 0,
  ifelse(gender == 0 & age == 14 & bmiscale <16.39, 0,
  ifelse(gender == 0 & age == 15 & bmiscale <16.98, 0,
  ifelse(gender == 0 & age == 16 & bmiscale <17.53, 0,
  ifelse(gender == 0 & age == 17 & bmiscale <18.04, 0,
  ifelse(gender == 1 & age == 2 & bmiscale <14.96, 0,
  ifelse(gender == 1 & age == 3 & bmiscale <14.60, 0,
  ifelse(gender == 1 & age == 4 & bmiscale <14.30, 0,
  ifelse(gender == 1 & age == 5 & bmiscale <14.04, 0,
  ifelse(gender == 1 & age == 6 & bmiscale <13.85, 0,
  ifelse(gender == 1 & age == 7 & bmiscale <13.83, 0,
  ifelse(gender == 1 & age == 8 & bmiscale <14.00, 0,
  ifelse(gender == 1 & age == 9 & bmiscale <14.26, 0,
  ifelse(gender == 1 & age == 10 & bmiscale <14.58, 0,
  ifelse(gender == 1 & age == 11 & bmiscale <15.03, 0,
  ifelse(gender == 1 & age == 12 & bmiscale <15.59, 0,
  ifelse(gender == 1 & age == 13 & bmiscale <16.23, 0,
  ifelse(gender == 1 & age == 14 & bmiscale <16.86, 0,
  ifelse(gender == 1 & age == 15 & bmiscale <17.43, 0,
  ifelse(gender == 1 & age == 16 & bmiscale <17.90, 0,
  ifelse(gender == 1 & age == 17 & bmiscale <18.24, 0,
  ifelse(age >= 18 & bmiscale <18.5, 0, NA))))))))))))))))))))))))))))))))))

HS14C <- HS14C %>%
  mutate(average =
  ifelse(gender == 0 & age == 2 & bmiscale >=15.24 & bmiscale <18.36, 1,
  ifelse(gender == 0 & age == 3 & bmiscale >=14.83 & bmiscale <17.85, 1,
  ifelse(gender == 0 & age == 4 & bmiscale >=14.51 & bmiscale <17.52, 1,
  ifelse(gender == 0 & age == 5 & bmiscale >=14.26 & bmiscale <17.39, 1,
  ifelse(gender == 0 & age == 6 & bmiscale >=14.06 & bmiscale <17.52, 1,
  ifelse(gender == 0 & age == 7 & bmiscale >=14.00 & bmiscale <17.88, 1,
  ifelse(gender == 0 & age == 8 & bmiscale >=14.13 & bmiscale <18.41, 1,
  ifelse(gender == 0 & age == 9 & bmiscale >=14.36 & bmiscale <19.07, 1,
  ifelse(gender == 0 & age == 10 & bmiscale >=14.63 & bmiscale <19.80, 1,
  ifelse(gender == 0 & age == 11 & bmiscale >=14.96 & bmiscale <20.51, 1,
  ifelse(gender == 0 & age == 12 & bmiscale >=15.36 & bmiscale <21.20, 1,
  ifelse(gender == 0 & age == 13 & bmiscale >=15.84 & bmiscale <21.89, 1,
```

```
  ifelse(gender == 0 & age == 14 & bmiscale >=16.39 & bmiscale <22.60, 1,
  ifelse(gender == 0 & age == 15 & bmiscale >=16.98 & bmiscale <23.28, 1,
  ifelse(gender == 0 & age == 16 & bmiscale >=17.53 & bmiscale <23.89, 1,
  ifelse(gender == 0 & age == 17 & bmiscale >=18.04 & bmiscale <24.46, 1,
  ifelse(gender == 1 & age == 2 & bmiscale >=14.96 & bmiscale <18.09, 1,
  ifelse(gender == 1 & age == 3 & bmiscale >=14.60 & bmiscale <17.64, 1,
  ifelse(gender == 1 & age == 4 & bmiscale >=14.30 & bmiscale <17.35, 1,
  ifelse(gender == 1 & age == 5 & bmiscale >=14.04 & bmiscale <17.23, 1,
  ifelse(gender == 1 & age == 6 & bmiscale >=13.85 & bmiscale <17.33, 1,
  ifelse(gender == 1 & age == 7 & bmiscale >=13.83 & bmiscale <17.69, 1,
  ifelse(gender == 1 & age == 8 & bmiscale >=14.00 & bmiscale <18.28, 1,
  ifelse(gender == 1 & age == 9 & bmiscale >=14.26 & bmiscale <18.99, 1,
  ifelse(gender == 1 & age == 10 & bmiscale >=14.58 & bmiscale <19.78, 1,
  ifelse(gender == 1 & age == 11 & bmiscale >=15.03 & bmiscale <20.66, 1,
  ifelse(gender == 1 & age == 12 & bmiscale >=15.59 & bmiscale <21.59, 1,
  ifelse(gender == 1 & age == 13 & bmiscale >=16.23 & bmiscale <22.49, 1,
  ifelse(gender == 1 & age == 14 & bmiscale >=16.86 & bmiscale <23.27, 1,
  ifelse(gender == 1 & age == 15 & bmiscale >=17.43 & bmiscale <23.89, 1,
  ifelse(gender == 1 & age == 16 & bmiscale >=17.90 & bmiscale <24.34, 1,
  ifelse(gender == 1 & age == 17 & bmiscale >=18.24 & bmiscale <24.70, 1,
  ifelse(age >= 18 & bmiscale >=18.5 & bmiscale <25, 1, NA)))))))))))))))))))))))))))))))))))

HS14C <- HS14C %>%
  mutate(over=
  ifelse(gender == 0 & age == 2 & bmiscale >=18.36 & bmiscale <19.99, 2,
  ifelse(gender == 0 & age == 3 & bmiscale >=17.85 & bmiscale <19.50, 2,
  ifelse(gender == 0 & age == 4 & bmiscale >=17.52 & bmiscale <19.23, 2,
  ifelse(gender == 0 & age == 5 & bmiscale >=17.39 & bmiscale <19.27, 2,
  ifelse(gender == 0 & age == 6 & bmiscale >=17.52 & bmiscale <19.76, 2,
  ifelse(gender == 0 & age == 7 & bmiscale >=17.88 & bmiscale <20.59, 2,
  ifelse(gender == 0 & age == 8 & bmiscale >=18.41 & bmiscale <21.56, 2,
  ifelse(gender == 0 & age == 9 & bmiscale >=19.07 & bmiscale <22.71, 2,
  ifelse(gender == 0 & age == 10 & bmiscale >=19.80 & bmiscale <23.96, 2,
  ifelse(gender == 0 & age == 11 & bmiscale >=20.51 & bmiscale <25.07, 2,
  ifelse(gender == 0 & age == 12 & bmiscale >=21.20 & bmiscale <26.02, 2,
  ifelse(gender == 0 & age == 13 & bmiscale >=21.89 & bmiscale <26.87, 2,
  ifelse(gender == 0 & age == 14 & bmiscale >=22.60 & bmiscale <27.64, 2,
  ifelse(gender == 0 & age == 15 & bmiscale >=23.28 & bmiscale <28.32, 2,
  ifelse(gender == 0 & age == 16 & bmiscale >=23.89 & bmiscale <28.89, 2,
  ifelse(gender == 0 & age == 17 & bmiscale >=24.46 & bmiscale <29.43, 2,
```

```
  ifelse(gender == 1 & age == 2 & bmiscale >=18.09 & bmiscale <19.81, 2,
  ifelse(gender == 1 & age == 3 & bmiscale >=17.64 & bmiscale <19.38, 2,
  ifelse(gender == 1 & age == 4 & bmiscale >=17.35 & bmiscale <19.16, 2,
  ifelse(gender == 1 & age == 5 & bmiscale >=17.23 & bmiscale <19.20, 2,
  ifelse(gender == 1 & age == 6 & bmiscale >=17.33 & bmiscale <19.61, 2,
  ifelse(gender == 1 & age == 7 & bmiscale >=17.69 & bmiscale <20.39, 2,
  ifelse(gender == 1 & age == 8 & bmiscale >=18.28 & bmiscale <21.44, 2,
  ifelse(gender == 1 & age == 9 & bmiscale >=18.99 & bmiscale <22.66, 2,
  ifelse(gender == 1 & age == 10 & bmiscale >=19.78 & bmiscale <23.97, 2,
  ifelse(gender == 1 & age == 11 & bmiscale >=20.66 & bmiscale <25.25, 2,
  ifelse(gender == 1 & age == 12 & bmiscale >=21.59 & bmiscale <26.47, 2,
  ifelse(gender == 1 & age == 13 & bmiscale >=22.49 & bmiscale <27.57, 2,
  ifelse(gender == 1 & age == 14 & bmiscale >=23.27 & bmiscale <28.42, 2,
  ifelse(gender == 1 & age == 15 & bmiscale >=23.89 & bmiscale <29.01, 2,
  ifelse(gender == 1 & age == 16 & bmiscale >=24.34 & bmiscale <29.40, 2,
  ifelse(gender == 1 & age == 17 & bmiscale >=24.70 & bmiscale <29.70, 2,
  ifelse(age >= 18 & bmiscale >=25 & bmiscale <30, 2, NA))))))))))))))))))))))))))))))))))

HS14C <- HS14C %>%
  mutate(obes=
  ifelse(gender == 0 & age == 2 & bmiscale >=19.99, 3,
  ifelse(gender == 0 & age == 3 & bmiscale >=19.50, 3,
  ifelse(gender == 0 & age == 4 & bmiscale >=19.23, 3,
  ifelse(gender == 0 & age == 5 & bmiscale >=19.27, 3,
  ifelse(gender == 0 & age == 6 & bmiscale >=19.76, 3,
  ifelse(gender == 0 & age == 7 & bmiscale >=20.59, 3,
  ifelse(gender == 0 & age == 8 & bmiscale >=21.56, 3,
  ifelse(gender == 0 & age == 9 & bmiscale >=22.71, 3,
  ifelse(gender == 0 & age == 10 & bmiscale >=23.96, 3,
  ifelse(gender == 0 & age == 11 & bmiscale >=25.07, 3,
  ifelse(gender == 0 & age == 12 & bmiscale >=26.02, 3,
  ifelse(gender == 0 & age == 13 & bmiscale >=26.87, 3,
  ifelse(gender == 0 & age == 14 & bmiscale >=27.64, 3,
  ifelse(gender == 0 & age == 15 & bmiscale >=28.32, 3,
  ifelse(gender == 0 & age == 16 & bmiscale >=28.89, 3,
  ifelse(gender == 0 & age == 17 & bmiscale >=29.43, 3,
  ifelse(gender == 1 & age == 2 & bmiscale >=19.81, 3,
  ifelse(gender == 1 & age == 3 & bmiscale >=19.38, 3,
  ifelse(gender == 1 & age == 4 & bmiscale >=19.16, 3,
  ifelse(gender == 1 & age == 5 & bmiscale >=19.20, 3,
```

```
  ifelse(gender == 1 & age == 6 & bmiscale >=19.61, 3,
  ifelse(gender == 1 & age == 7 & bmiscale >=20.39, 3,
  ifelse(gender == 1 & age == 8 & bmiscale >=21.44, 3,
  ifelse(gender == 1 & age == 9 & bmiscale >=22.66, 3,
  ifelse(gender == 1 & age == 10 & bmiscale >=23.97, 3,
  ifelse(gender == 1 & age == 11 & bmiscale >=25.25, 3,
  ifelse(gender == 1 & age == 12 & bmiscale >=26.47, 3,
  ifelse(gender == 1 & age == 13 & bmiscale >=27.57, 3,
  ifelse(gender == 1 & age == 14 & bmiscale >=28.42, 3,
  ifelse(gender == 1 & age == 15 & bmiscale >=29.01, 3,
  ifelse(gender == 1 & age == 16 & bmiscale >=29.40, 3,
  ifelse(gender == 1 & age == 17 & bmiscale >=29.70, 3,
  ifelse(age >= 18 & bmiscale >=30, 3, NA))))))))))))))))))))))))))))))))))


# 0: underweight, 1: average, 2: overweight, 3: obese
HS14C <- HS14C %>%
  mutate(bmic = ifelse(under == 0 & is.na(average) & is.na(over) & is.na(obes), 0,
                ifelse(average == 1 & is.na(under) & is.na(over) & is.na(obes), 1,
                ifelse(over == 2 & is.na(under) & is.na(average) & is.na(obes), 2,
                ifelse(obes == 3 & is.na(under) & is.na(over) & is.na(average), 3, NA)))))

HS14C$bmic <- as.factor(HS14C$bmic)
summary(HS14C$bmic)

# exclude underweight
c14<- subset(HS14C, !bmic == 0 | is.na(bmic))


## BMI
summary(c14$bmiscale)

c14$bmic <- factor(c14$bmic, labels=c("Average","Overweight","Obese"))

prop_frame(c14, "bmic")


## Sex
c14$gender <- factor(c14$gender, labels=c("Male","Female"))

```

```r
prop_frame(c14, "gender")


## Age
summary(c14$age)


## Ethnicity
# recode anyone ~ european
c14 = c14 %>%
  mutate(euro = ifelse(C4_03_01 == 1, 1, 0))

# recode anyone ~ maori
c14 = c14 %>%
  mutate(maori = ifelse(C4_03_02 == 1, 1, 0))

# recode anyone ~ pacific
c14 = c14 %>%
  mutate(pacific = ifelse(C4_03_03 == 1, 1,
                   ifelse(C4_03_04 == 1, 1,
                   ifelse(C4_03_05 == 1, 1,
                   ifelse(C4_03_06 == 1, 1, 0)))))

# recode anyone ~ asian (Indian and Chinese)
c14 = c14 %>%
  mutate(asian = ifelse(C4_03_07 == 1, 1,
                  ifelse(C4_03_08 == 1, 1, 0)))

# recode anyone ~ other
c14 = c14 %>%
  mutate(other = ifelse(C4_03_77 == 1, 1, 0))


c14 = c14 %>%
  mutate(eth_count = ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "Maori
      Only",
                     ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "
                        Pacific Only",
                     ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Asian
                        Only",
```

```
                ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 0, "European Only",
                ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other",
                ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+ Ethnicities (M)",
                ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 0, "2+ Ethnicities (M)",
                ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 0, "Other",
                ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 0, "Other",
                ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 0 & other == 1, "Other",
                ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other",
                ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other",
                ifelse(euro == 0 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other",
                ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 0, "2+ Ethnicities (M)",
                ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 0 & other == 1, "2+ Ethnicities (M)",
                ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 0, "Other",
                ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 0 & other == 1, "Other",
                ifelse(euro == 1 & maori == 0 & pacific == 0 & asian == 1 & other == 1, "Other",
                ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+ Ethnicities (M)",
                ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+ Ethnicities (M)",
```

```
                        ifelse(euro == 0 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+
                            Ethnicities (M)",
                        ifelse(euro == 0 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other"
                            ,
                        ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 0, "2+
                            Ethnicities (M)",
                        ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 0 & other == 1, "2+
                            Ethnicities (M)",
                        ifelse(euro == 1 & maori == 1 & pacific == 0 & asian == 1 & other == 1, "2+
                            Ethnicities (M)",
                        ifelse(euro == 1 & maori == 0 & pacific == 1 & asian == 1 & other == 1, "Other"
                            ,
                        ifelse(euro == 0 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
                            Ethnicities (M)",
                        ifelse(euro == 1 & maori == 1 & pacific == 1 & asian == 1 & other == 1, "2+
                            Ethnicities (M)", NA)))))))))))))))))))))))))))))))

c14$eth_count <- as.factor(c14$eth_count)

prop_frame(c14, "eth_count")


## Household Income
# combine adult to child data to match income and education data
ac14 <- left_join(select(c14, HHID, C4_17, C4_20, C4_21),
                  select(a14, HHID, A5_24, A5_14, A5_15), by="HHID")

summary(ac14$A5_24)
ac14$hhinc1 <- ordered(ac14$A5_24, levels=c("4","5","6","7","8","9","10","11","12",
                                            "13","14","15","16"))
summary(ac14$C4_17)
ac14$hhinc2 <- ordered(ac14$C4_17, levels=c("4","5","6","7","8","9","10","11","12",
                                            "13","14","15","16"))

ac14$hhinc1 <- as.numeric(as.character(ac14$hhinc1))
ac14$hhinc2 <- as.numeric(as.character(ac14$hhinc2))

ac14 <- mutate(ac14, inc= ifelse(is.na(hhinc2), hhinc1,
                             ifelse(is.na(hhinc1), hhinc2, hhinc1)))

```

```
ac14$hhinc <- cut(ac14$inc, breaks=c(0,5,10,13,16))
ac14$hhinc <- factor(ac14$hhinc, labels=c("<=$15,000","$15,001-$40,000","$40,001-$70,000",">$70,000"
    ))

ac14$hhinc <- as.factor(ac14$hhinc)
summary(ac14$hhinc)

# recode back into original data set
c14$hhinc <- ac14$hhinc

prop_frame(c14, "hhinc")


## Education
ac14$cedu1 <- as.numeric(as.character(ac14$C4_20))
ac14$aedu1 <- as.numeric(as.character(ac14$A5_14))

ac14 <- mutate(ac14, edu1 = ifelse(is.na(cedu1), aedu1,
                             ifelse(is.na(aedu1), cedu1,
                             ifelse(cedu1 >= aedu1, cedu1,
                             ifelse(aedu1 >= cedu1, aedu1, NA)))))

ac14$edu1 <- as.factor(ac14$edu1)

ac14$secondary <- as.numeric(as.character(ac14$edu1))
ac14$secondary <- recode(ac14$secondary, "1" = "1", .default = "2") # 1= no degree, 2= secondary
    school grads
ac14$secondary <- as.factor(ac14$secondary)
summary(ac14$secondary)


ac14$C4_21[ac14$C4_21 == "77"] <- NA
ac14$A5_15[ac14$A5_15 == "77"] <- NA

ac14$cedu2 <- as.numeric(as.character(ac14$C4_21))
ac14$aedu2 <- as.numeric(as.character(ac14$A5_15))

ac14 <-  mutate(ac14, edu2 = ifelse(is.na(cedu2), aedu2,
                             ifelse(is.na(aedu2), cedu2,
                             ifelse(cedu2 >= aedu2, cedu2,
```

```r
                                ifelse(aedu2 >= cedu2, aedu2, NA)))))

ac14$edu2 <- as.factor(ac14$edu2)

ac14$tertiary <- as.numeric(as.character(ac14$edu2))
ac14$tertiary <- recode(ac14$tertiary, "0" = "0","1"="2","2"="2",
                          "3"="2", "4"="2", .default = "3") #0 = no tertiary, 3= tertiary
ac14$tertiary <- as.factor(ac14$tertiary)

# recode into one variable, prioritise highest education
ac14 <- mutate(ac14, edu = ifelse(secondary =="1" & tertiary =="0", 1,
                            ifelse(secondary == "2" & tertiary == "0", 2,
                            ifelse(tertiary == "2", 2,
                            ifelse(tertiary == "3", 3, 1)))))
summary(ac14$edu)

# recode back into child data set
c14$edu <- factor(ac14$edu, labels=c("No Qualification","Secondary","Tertiary"))

prop_frame(c14, "edu")

# remove ac14 from memory
rm(ac14)


## Deprivation Quintile
names(c14)[names(c14) == "nzdep13_quintile"] <- "dep"

prop_frame(c14,"dep")


## Fruit and Vegetable
c14$kidfruit_2serves[c14$kidfruit_2serves == "K"] <- NA
c14$kidfruit_2serves[c14$kidfruit_2serves == "X"] <- NA
c14$kidveges_new[c14$kidveges_new == "K"] <- NA
c14$kidveges_new[c14$kidveges_new == "R"] <- NA
c14$kidveges_new[c14$kidveges_new == "X"] <- NA

c14$fruit <- factor(c14$kidfruit_2serves, labels=c("No","Yes"))
c14$veges <- factor(c14$kidveges_new, labels=c("No","Yes"))
```

```
prop_frame(c14, "fruit")
prop_frame(c14, "veges")


## Urban/Rural Area
c14$UA <- recode(c14$UrbanArea2015Class, "M"="Urban","S"="Urban",.default="Rural")

prop_frame(c14, "UA")


## Soft Drink and Fast Food
summary(c14$C3_09) # 21 K
c14$softd <- as.numeric(as.character(c14$C3_09))
c14$softd <- cut(c14$softd, c(0,1,2,4,Inf),labels=c("0/week","1/week",
                  "2-3/week","4+/week"), right=FALSE)

summary(c14$C3_10) # 11 K
c14$fastf <- as.numeric(as.character(c14$C3_10))
c14$fastf <- cut(c14$fastf, c(0,1,2,4,Inf),labels=c("0/week","1/week",
                  "2-3/week","4+/week"), right=FALSE)

prop_frame(c14, "softd")
prop_frame(c14, "fastf")
```

Listing A.2: The NZHSs Data Analyses

```
#### Data Analyses – New Zealand Health Surveys

#### Merge the NZHSs together ####
# Merge adult and child data, then split them into two groups: 2–17 years old and 18+ years old
    group
# 2002/03 nzhs does not have a separate child data
ac06 <- merge(a06, c06, all=TRUE)
ac11 <- merge(a11, c11, all=TRUE)
ac12 <- merge(a12, c12, all=TRUE)
ac13 <- merge(a13, c13, all=TRUE)
ac14 <- merge(a14, c14, all=TRUE)

# add year variable to all data sets
hs02$year <- rep("2002/03")
ac06$year <- rep("2006/07")
ac11$year <- rep("2011/12")
ac12$year <- rep("2012/13")
ac13$year <- rep("2013/14")
ac14$year <- rep("2014/15")

# Split them into child data (2--17 years old) and adult data (18+ years old)
adult <- subset(hs02, age >=18) %>%
         merge(subset(ac06, age >=18), all=TRUE) %>%
         merge(subset(ac11, age >=18), all=TRUE) %>%
         merge(subset(ac12, age >=18), all=TRUE) %>%
         merge(subset(ac13, age >=18), all=TRUE) %>%
         merge(subset(ac14, age >=18), all=TRUE)

child <- subset(ac06, age >=2 & age <18) %>%
         merge(subset(ac11, age >=2 & age <18), all=TRUE) %>%
         merge(subset(ac12, age >=2 & age <18), all=TRUE) %>%
         merge(subset(ac13, age >=2 & age <18), all=TRUE) %>%
         merge(subset(ac14, age >=2 & age <18), all=TRUE)



#### Missing Data analysis ####

```

```r
## Missing BMI data
# adults
adult %>%
  group_by(year) %>%
  summarise(n=n(), nmiss=sum(is.na(bmiscale))) %>%
  mutate(prop.miss = nmiss/n)

# children
child %>%
  group_by(year) %>%
  summarise(n=n(), nmiss=sum(is.na(bmic))) %>%
  mutate(prop.miss = nmiss/n)


## Missing Household Income
# adults
adult %>%
  group_by(year) %>%
  summarise(n=n(), nmiss=sum(is.na(hhinc))) %>%
  mutate(prop.miss = nmiss/n)

# children
child %>%
  group_by(year) %>%
  summarise(n=n(), nmiss=sum(is.na(hhinc))) %>%
  mutate(prop.miss = nmiss/n)


## Missing Education Data
# adults
adult %>%
  group_by(year) %>%
  summarise(n=n(), nmiss=sum(is.na(edu))) %>%
  mutate(prop.miss = nmiss/n)

# children
child %>%
  group_by(year) %>%
  summarise(n=n(), nmiss=sum(is.na(edu))) %>%
  mutate(prop.miss = nmiss/n)
```

```
#### Histogram, exclude missing BMI data, include relevant variables ####

# histogram and qq plot
normal_plot(adult, "bmiscale")
# not normally distributed

# BMI histogram across ethnicities
h_eu = ggplot(subset(adult, eth_count == "European Only"), aes(bmiscale)) +
  geom_histogram() +
  labs(title="European Only") +
  scale_x_continuous(limits=c(18,80), breaks=seq(0,100,10))

h_maori = ggplot(subset(adult, eth_count == "Maori Only"), aes(bmiscale)) +
  geom_histogram() +
  labs(title="Maori Only") +
  scale_x_continuous(limits=c(18,80), breaks=seq(0,100,10))

h_pa = ggplot(subset(adult, eth_count == "Pacific Only"), aes(bmiscale)) +
  geom_histogram() +
  labs(title="Pacific Only") +
  scale_x_continuous(limits=c(18,80), breaks=seq(0,100,10))

h_asian = ggplot(subset(adult, eth_count == "Asian Only"), aes(bmiscale)) +
  geom_histogram() +
  labs(title="Asian Only") +
  scale_x_continuous(limits=c(18,80), breaks=seq(0,100,10))

h_2 = ggplot(subset(adult, eth_count == "2+ Ethnicities (M)"), aes(bmiscale)) +
  geom_histogram() +
  labs(title="2+ Ethnicities (M)") +
  scale_x_continuous(limits=c(18,80), breaks=seq(0,100,10))

h_other = ggplot(subset(adult, eth_count == "Other"), aes(bmiscale)) +
  geom_histogram() +
  labs(title="Other") +
  scale_x_continuous(limits=c(18,80), breaks=seq(0,100,10))

```

```r
grid.arrange(h_eu,h_maori,h_pa,h_asian,h_2,h_other, nrow=6, ncol=1)



# measure skewness and kurtosis for bmiscale (numeric)
skewness(adult$bmiscale, na.rm=TRUE)
kurtosis(adult$bmiscale, na.rm=TRUE)

# Agostino test (H0 = skewness equal to zero)
agostino.test(adult$bmiscale, alternative="less") # sample size is too large

# Anscombe-Glynn test (H0 = kurtosis equal to 3)
anscombe.test(adult$bmiscale, alternative="less")


# Subset the Adult data to exclude missing BMI value
adult <- subset(adult, !is.na(bmiscale))


# Transform them into tertiles
adult$bmi_t <- factor(ntile(adult$bmiscale, 3),
                        labels=c("Bottom Tertile","Mid Tertile","Top Tertile"), ordered=TRUE)


# Identify the BMI range on each tertile
adult %>%
  group_by(bmi_t) %>%
  summarise_each(funs(min, max), bmiscale)


## Subset the data to only include relevant variables

adult <- subset(adult, select = c("age","bmic","gender","bmiscale", "UA", "bmi_t",
                                  "dep", "hhinc","edu", "active", "native",
                                  "sedentary","stair","fruit","veges",
                                  "year","smoke", "eth_count", "finalwgt", "cluster", "strata",
                                  "haz_drinker_all"))

# also exclude missing BMI value in the Child data
child <- subset(child, !is.na(bmic))
```

```r

child <- subset(child, select=c("age","bmic","gender","bmiscale", "UA",
                                "dep", "hhinc","edu","fruit","veges",
                                "year","softd","fastf", "eth_count",
                                "finalwgt", "strata", "cluster"))



#### Bivariate Analysis - Adult ####
# All possible combination of explanatory variables with the outcome variable
# ANOVA = numeric x categorical; chi-squared = categorical x categorical
# We did not use sample weights in the following analyses

## Unweigthed Stacked plots
# Age X bmi_t (BMI tertile)
aage = box_notitle(adult, "bmi_t", "age", "BMI Tertile", "Age", min = 15, gap = 5)
summary(lm(age ~ bmi_t, data = adult))
anova_table(adult$age, adult$bmi_t)

t = aggregate(age ~ bmi_t, adult, function(x) c(mean = mean(x), sd = sd(x)))


# a function to copy tables to excel
write.excel <- function(x,row.names=TRUE,col.names=TRUE,...) {
  write.table(x,"clipboard",sep="\t",row.names=row.names,col.names=col.names,...)
}
# copied the object "t" and moved them to excel each time

# Sex
asex = stacked_notitle(adult, "bmi_t", "gender", "Sex", "BMI Tertiles")
t = table(adult$bmi_t, adult$gender)
t = round(prop.table(table(adult$bmi_t, adult$gender), 1),2)

# Ethnicity
aeth = stacked_notitle(adult, "bmi_t", "eth_count", "Ethnicity", "BMI Tertile")
t = table(adult$bmi_t, adult$eth_count)
t = prop.table(table(adult$bmi_t, adult$eth_count), 1)

# Deprivation quintile
adep = stacked_notitle(adult, "bmi_t", "dep", "Deprivation Quintile", "BMI Tertile")
```

```r
t = table(adult$bmi_t, adult$dep)
t = prop.table(table(adult$bmi_t, adult$dep), 1)

# Urban/Rural
aua = stacked_notitle(adult, "bmi_t", "UA", "Urban/Rural Area", "BMI Tertile")
t = table(adult$bmi_t, adult$UA)
t = prop.table(table(adult$bmi_t, adult$UA), 1)

# Household Income
ainc = stacked_notitle(adult, "bmi_t", "hhinc", "Household Income", "BMI Tertile")
t = table(adult$bmi_t, adult$hhinc)
t = prop.table(table(adult$bmi_t, adult$hhinc), 1)

# Education
aedu = stacked_notitle(adult, "bmi_t", "edu", "Education", "BMI Tertile")
t = table(adult$bmi_t, adult$edu)
t = prop.table(table(adult$bmi_t, adult$edu), 1)

# Fruit Consumption
afruit = stacked_notitle(adult, "bmi_t", "fruit", "Fruit Guideline", "BMI Tertile")
t = table(adult$bmi_t, adult$fruit)
t = prop.table(table(adult$bmi_t, adult$fruit), 1)

# Vegetable Consumption
avege = stacked_notitle(adult, "bmi_t", "veges", "Vegetable Guideline", "BMI Tertile")
t = table(adult$bmi_t, adult$veges)
t = prop.table(table(adult$bmi_t, adult$veges), 1)


# Migration Status
amigrat = stacked_notitle(adult, "bmi_t", "native", "Migration Status", "BMI Tertile")
t = table(adult$bmi_t, adult$native)
t = prop.table(table(adult$bmi_t, adult$native), 1)

# Difficulty Climbing Stairs
astair = stacked_notitle(adult, "bmi_t", "stair", "Difficulty Climbing \nSeveral Flights \nof Stairs
    ", "BMI Tertile")
t = table(adult$bmi_t, adult$stair)
t = prop.table(table(adult$bmi_t, adult$stair), 1)

```

```
# Smoking Status
asmoke = stacked_notitle(adult, "bmi_t", "smoke", "Smoking Status", "BMI Tertile")
t = table(adult$bmi_t, adult$smoke)
t = prop.table(table(adult$bmi_t, adult$smoke), 1)

# Alcohol Problem
aalc = stacked_notitle(adult, "bmi_t", "haz_drinker_all", "Drinking Problem", "BMI Tertile")
t = table(adult$bmi_t, adult$haz_drinker_all)
t = prop.table(table(adult$bmi_t, adult$haz_drinker_all), 1)

# Physically Active
aact = stacked_notitle(adult, "bmi_t", "active", "Physical Activity", "BMI Tertile")
t = table(adult$bmi_t, adult$active)
t = prop.table(table(adult$bmi_t, adult$active), 1)

# Sedentary Lifestyle
aseden = stacked_notitle(adult, "bmi_t", "sedentary", "Sedentary Lifestyle", "BMI Tertile")
t = table(adult$bmi_t, adult$sedentary)
t = prop.table(table(adult$bmi_t, adult$sedentary), 1)


# arrange all the charts
grid.arrange(aage, asex, aeth, adep, aua, ainc, aedu, afruit, avege, amigrat, astair, asmoke, aalc,
    aact, aseden,
             ncol = 3, nrow = 5, padding = unit(3, "line"))


## Chi-Squared tests (unweighted)
label = c("Sex", "Ethnicity", "Deprivation Quintile", "Urban/Rural Area", "Household Income","
    Education","Fruit Guideline",
          "Vegetable Guideline","Migration Status","Difficulty Climbing Several Flights of Stairs","
              Smoking Status","Alcohol Problem",
          "Physical Activity","Sedentary Lifestyle")

table = rbind(chix_table(adult$bmi_t, adult$gender),chix_table(adult$bmi_t, adult$eth_count),chix_
    table(adult$bmi_t, adult$dep),
      chix_table(adult$bmi_t, adult$UA),chix_table(adult$bmi_t, adult$hhinc),chix_table(adult$bmi_t,
          adult$edu),
      chix_table(adult$bmi_t, adult$fruit),chix_table(adult$bmi_t, adult$veges),chix_table(adult$bmi
          _t, adult$native),
```

```r
    chix_table(adult$bmi_t, adult$stair),chix_table(adult$bmi_t, adult$smoke),chix_table(adult$bmi_t, adult$haz_drinker_all),
    chix_table(adult$bmi_t, adult$active),chix_table(adult$bmi_t, adult$sedentary))

biv_table_adult = cbind(label, table)

# remove temporary object from memory
rm(label)
rm(table)


#### Relevel and set reference level - Adult ####

str(adult)

# relevel, set reference level
adult$eth_count = as.factor(adult$eth_count)
adult$eth_count <- relevel(adult$eth_count, ref="European Only")
adult$gender <- relevel(adult$gender, ref="Male")
adult$dep = relevel(adult$dep, ref=1)
adult$hhinc = relevel(adult$hhinc, ref="<=$15,000")
adult$hhinc = droplevels(adult$hhinc)
adult$edu = relevel(adult$edu, ref="No Qualification")
adult$stair = factor(adult$stair, levels = c("No Difficulty","A Little Difficult","A Lot Difficult")
  )
adult$stair = relevel(adult$stair, ref="No Difficulty")
adult$sedentary = as.factor(adult$sedentary)
adult$sedentary = relevel(adult$sedentary, ref="Not Sedentary")
adult$smoke = relevel(adult$smoke, ref="Non Smoker")
adult$veges = droplevels(adult$veges)
adult$veges = relevel(adult$veges, ref="No")
adult$fruit = relevel(adult$fruit, ref="No")
adult$active = relevel(adult$active, ref="Active")
adult$year = as.factor(adult$year)
adult$year = relevel(adult$year, ref="2014/15")
adult$bmic = ordered(adult$bmic)
adult$native = droplevels(adult$native)
adult$native = relevel(adult$native, ref="Migrant")
adult$haz_drinker_all = droplevels(adult$haz_drinker_all)

```

```r


#### Survey Design Object - Adult ####

adult$finalwgt <- as.numeric(adult$finalwgt)
summary(adult$finalwgt) # 1 missing value

# exclude the datum with missing cluster them from the analysis
adult <- subset(adult, !is.na(finalwgt))

# creating a linear contrast
contr.lin <- contr.poly(6)[,1]

# bind the contrast to the data set
adult$contr.lin <- contr.lin[ match (adult$year, c("2002/03","2006/07", "2011/12",
                                                   "2012/13","2013/14", "2014/15"))]

# Create a complex survey design object for the adult data
a.design <- svydesign(
  id = ~cluster,
  strata = ~interaction(strata,year),
  data = adult,
  weights = ~finalwgt,
  nest = TRUE
)

print(summary(a.design))


#### BMI Category by Year - Adult ####


## stacked plot for BMI categories by year (non age-standardised, using survey weights)
a.bmi_plot = svyby(~bmic, ~year, svymean, design=a.design, vartype = c('ci','se'))
a.bmi_plot = data.frame(a.bmi_plot)
a.bmi_plot <- gather(a.bmi_plot, bmic, prop, 2:4, factor_key = TRUE)

a.bmi_plot = a.bmi_plot %>%
  group_by(year) %>%
  mutate(pos = cumsum(prop) - 0.5*prop) # calculating position
```

```
a.bmi_plot$label_p = paste0(sprintf("%.1f", a.bmi_plot$prop*100),"%")

bmic_adult_year = ggplot(a.bmi_plot, aes(x=year, y=prop, fill=bmic, levels=bmic)) +
  geom_bar(stat="identity", width = 0.8) +
  labs(title= "Adult Data", x= "Year", y= "Density") +
  scale_fill_discrete(guide=FALSE) +
  theme(axis.title = element_text(size = 12, face="bold"),
        plot.title = element_text(size = 14, face="bold", hjust=0.5),
        legend.title = element_text(size = 12, face="bold"),
        axis.text.x = element_text(size = 11),
        axis.text.y = element_text(size = 11)) +
  geom_text(aes(label=label_p), position = position_stack(vjust = 0.5), size=4, colour = "white") +
  scale_x_discrete(limits=c("2002/03","2006/07","2011/12","2012/13","2013/14","2014/15"),
                   labels=c("'02/03","'06/07","'11/12","'12/13","'13/14", "'14/15"))


## age-standardised prevalence
adult$agegrp = cut(adult$age, breaks = c(seq(0,85, by = 5), Inf), right = FALSE)

obes_a = adult %>%
  group_by(year, agegrp, eth_count, bmic) %>%
  filter(!is.na(eth_count)) %>%
  summarise(n=n())

obes_a = spread(obes_a, bmic, n)

# replace NA with 0
obes_a[is.na(obes_a)] = 0

obes_a$total = obes_a$Average + obes_a$Overweight + obes_a$Obese

obes_a$age_aver = (obes_a$Average/obes_a$total)*100
obes_a$age_over = (obes_a$Overweight/obes_a$total)*100
obes_a$age_obes = (obes_a$Obese/obes_a$total)*100

# scandinavian pop proportion
scand_prop_adult <- c(rep(0.07, 8), 0.06, 0.05, 0.04, 0.03, 0.02, 0.01, 0.01)
# rescale them so that they add up to 100%
scand_prop_adult <- scand_prop_adult*(1/((8*0.07)+0.06+0.05+0.04+0.03+0.02+0.01+0.01))
```

```
obes_a$scand_prop_adult = scand_prop_adult[ match (obes_a$agegrp, c("[15,20)", "[20,25)",
                                        "[25,30)", "[30,35)", "[35,40)", "[40,45)", "[45,50)",
                                        "[50,55)", "[55,60)", "[60,65)", "[65,70)", "[70,75)",
                                        "[75,80)", "[80,85)", "[85,Inf)"))]

obes_a$aver_rate = obes_a$age_aver*obes_a$scand_prop_adult
obes_a$over_rate = obes_a$age_over*obes_a$scand_prop_adult
obes_a$obes_rate = obes_a$age_obes*obes_a$scand_prop_adult

# add them up
obes_a1 = aggregate(cbind(aver_rate,over_rate,obes_rate) ~ year+eth_count, FUN = sum, data = obes_a,
          na.rm = TRUE, na.action = NULL)

obes_a1$over_obes = obes_a1$over_rate+obes_a1$obes_rate

# plot the obesity prevalence by ethnicity
aobes_plot = ggplot(obes_a1, aes(x=year, y=obes_rate, shape=eth_count, group=eth_count), stat="
    identity") +
  geom_line(size=1) +
  geom_point(size=2) +
  theme_light() +
  scale_x_discrete(limits=c("2002/03","2006/07","2011/12","2012/13","2013/14","2014/15")) +
  scale_y_continuous(limits=c(0,80),breaks=seq(0,100,10)) +
  labs(title="Adult Data", x = "Year", y = "Percentage", shape="Ethnicity") +
  theme(axis.title = element_text(size = 14, face="bold"),
        legend.title = element_text(size = 14, face="bold"),
        legend.text = element_text(size = 12),
        axis.text.x = element_text(size = 12),
        axis.text.y = element_text(size = 12),
        legend.position="none",
        plot.title = element_text(size = 16, face="bold", hjust=0.5))


# plot the over/obes prev
aovob_plot = ggplot(obes_a1, aes(x=year, y=over_rate+obes_rate, shape=eth_count, group=eth_count),
    stat="identity") +
  geom_line(size=1) +
  geom_point(size=2) +
  theme_light() +
```

```
  scale_x_discrete(limits=c("2002/03","2006/07","2011/12","2012/13","2013/14","2014/15")) +
  scale_y_continuous(limits=c(20,100),breaks=seq(0,100,10)) +
  labs(title="Adult Data", x = "Year", y = "Percentage", shape="Ethnicity") +
  theme(axis.title = element_text(size = 14, face="bold"),
        legend.title = element_text(size = 14, face="bold"),
        legend.text = element_text(size = 12),
        axis.text.x = element_text(size = 12),
        axis.text.y = element_text(size = 12),
        legend.position="none",
        plot.title = element_text(size = 16, face="bold", hjust=0.5))




#### Descriptive analysis across survey periods – Adult ####

# total sample size on each year
adult %>%
  group_by(year) %>%
  summarise(row = length(age))

# age summary
svyquantile(~age, subset(a.design, year == "2002/03"), c(0,.25,.5,.75,1), ci=TRUE)
svyquantile(~age, subset(a.design, year == "2006/07"), c(0,.25,.5,.75,1), ci=TRUE)
svyquantile(~age, subset(a.design, year == "2011/12"), c(0,.25,.5,.75,1), ci=TRUE)
svyquantile(~age, subset(a.design, year == "2012/13"), c(0,.25,.5,.75,1), ci=TRUE)
svyquantile(~age, subset(a.design, year == "2013/14"), c(0,.25,.5,.75,1), ci=TRUE)
svyquantile(~age, subset(a.design, year == "2014/15"), c(0,.25,.5,.75,1), ci=TRUE)

# bmi summary
svyquantile(~bmiscale, subset(a.design, year == "2002/03"), c(0,.25,.5,.75,1), ci=TRUE)
svyquantile(~bmiscale, subset(a.design, year == "2006/07"), c(0,.25,.5,.75,1), ci=TRUE)
svyquantile(~bmiscale, subset(a.design, year == "2011/12"), c(0,.25,.5,.75,1), ci=TRUE)
svyquantile(~bmiscale, subset(a.design, year == "2012/13"), c(0,.25,.5,.75,1), ci=TRUE)
svyquantile(~bmiscale, subset(a.design, year == "2013/14"), c(0,.25,.5,.75,1), ci=TRUE)
svyquantile(~bmiscale, subset(a.design, year == "2014/15"), c(0,.25,.5,.75,1), ci=TRUE)


# categorical variables
svyby(~bmic, ~year, svymean, design=a.design)
```

```
svyby(~gender, ~year, svymean, design=a.design)
svyby(~eth_count, ~year, svymean, design=a.design)
svyby(~dep, ~year, svymean, design=a.design)
svyby(~UA, ~year, svymean, design=a.design)
svyby(~hhinc, ~year, svymean, design=a.design)
svyby(~edu, ~year, svymean, design=a.design)
svyby(~fruit, ~year, svymean, design=a.design)
svyby(~veges, ~year, svymean, design=a.design)
svyby(~native, ~year, svymean, design=a.design)
svyby(~stair, ~year, svymean, design=a.design)
svyby(~smoke, ~year, svymean, design=a.design)
svyby(~haz_drinker_all, ~year, svymean, design=a.design)
svyby(~active, svymean, ~year, design=a.design)
svyby(~sedentary, ~year, svymean, design=a.design)



#### Trends over the survey periods - Adult ####
# reference level = 2014/15
summy(svyglm(age ~ year, a.design))
summy(svyglm(I(bmic == "Obese") ~ year, a.design))
summy(svyglm(I(gender == "Male" ) ~ year, a.design))
summy(svyglm(I(eth_count == "European Only") ~ year, a.design))
summy(svyglm(I(eth_count == "Other") ~ year, a.design))
summy(svyglm(I(UA == "Urban" ) ~ year, a.design))
summy(svyglm(I(fruit == "Yes") ~ year, a.design))
summy(svyglm(I(veges == "Yes") ~ year, a.design))
summy(svyglm(I(active == "Active") ~ year, a.design))
summy(svyglm(I(sedentary == "Sedentary") ~ year, a.design))


#### Forward Step Proportional Odds Logistic Regression - Adult ####
# Add one covariate at a time, add the strongest covariate on each step, and remove any non
    significant variable (p < 0.1)

## 1st step
# add deprivation and linear changes, variables of interest
olr1_dep = svyolr(bmi_t ~ dep, a.design)
summy(olr1_dep)

```

```
olr1_dep.lin = svyolr(bmi_t ~ dep + contr.lin, a.design)
summy(olr1_dep.lin)

# add other explanatory variables and try to find the strongest effect to add into the next step
olr1_inc = svyolr(bmi_t ~ hhinc + dep + contr.lin, a.design)
olr1_vege = svyolr(bmi_t ~ veges + dep + contr.lin, a.design)
olr1_age = svyolr(bmi_t ~ age + dep + contr.lin, a.design)
olr1_sex = svyolr(bmi_t ~ gender + dep + contr.lin, a.design)
olr1_eth = svyolr(bmi_t ~ eth_count + dep + contr.lin, a.design)
olr1_edu = svyolr(bmi_t ~ edu + dep + contr.lin, a.design)
olr1_seden = svyolr(bmi_t ~ sedentary + dep + contr.lin, a.design)
olr1_act = svyolr(bmi_t ~ active + dep + contr.lin, a.design)
olr1_stair = svyolr(bmi_t ~ stair + dep + contr.lin, a.design)
olr1_smoke = svyolr(bmi_t ~ smoke + dep + contr.lin, a.design)
olr1_fruit = svyolr(bmi_t ~ fruit + dep + contr.lin, a.design)
olr1_native = svyolr(bmi_t ~ native + dep + contr.lin, a.design)
olr1_alc = svyolr(bmi_t ~ haz_drinker_all + dep + contr.lin, a.design)
olr1_UA = svyolr(bmi_t ~ UA + dep + contr.lin, a.design)

summy(olr1_inc)
summy(olr1_vege)
summy(olr1_age)
summy(olr1_sex)
summy(olr1_eth)
summy(olr1_edu)
summy(olr1_seden)
summy(olr1_act)
summy(olr1_stair)
summy(olr1_smoke)
summy(olr1_fruit)
summy(olr1_native)
summy(olr1_alc)
summy(olr1_UA)
# alcohol problem and adherence to vegetable quideline were insignificant
# Ethnicity had the strongest effect


## 2nd step (bmi_t ~ ... + dep + contr.lin, a.design)
olr2_1 = svyolr(bmi_t ~ age + eth_count + dep + contr.lin, a.design)
olr2_2 = svyolr(bmi_t ~ gender + eth_count + dep + contr.lin, a.design)
```

```
olr2_3 = svyolr(bmi_t ~ edu + eth_count + dep + contr.lin , a.design)
olr2_4 = svyolr(bmi_t ~ sedentary + eth_count + dep + contr.lin , a.design)
olr2_5 = svyolr(bmi_t ~ stair + eth_count + dep + contr.lin , a.design)
olr2_6 = svyolr(bmi_t ~ smoke + eth_count + dep + contr.lin , a.design)
olr2_7 = svyolr(bmi_t ~ age + hhinc + dep + contr.lin , a.design)
olr2_8 = svyolr(bmi_t ~ active + eth_count + dep + contr.lin , a.design)
olr2_9 = svyolr(bmi_t ~ native + eth_count + dep + contr.lin , a.design)
olr2_10 = svyolr(bmi_t ~ UA + eth_count + dep + contr.lin , a.design)
olr2_11 = svyolr(bmi_t ~ hhinc + eth_count + dep + contr.lin , a.design)
olr2_12 = svyolr(bmi_t ~ fruit + eth_count + dep + contr.lin , a.design)

summy(olr2_1)
summy(olr2_2)
summy(olr2_3)
summy(olr2_4)
summy(olr2_5)
summy(olr2_6)
summy(olr2_7)
summy(olr2_8)
summy(olr2_9)
summy(olr2_10)
summy(olr2_11)
summy(olr2_12)
# adherence to fruit guideline was insignificant
# Age had the strongest effect


## 3rd step
olr3_1 = svyolr(bmi_t ~ gender + age + eth_count + dep + contr.lin , a.design)
olr3_2 = svyolr(bmi_t ~ smoke + age + eth_count + dep + contr.lin , a.design)
olr3_3 = svyolr(bmi_t ~ sedentary + age + eth_count + dep + contr.lin , a.design)
olr3_4 = svyolr(bmi_t ~ edu + age + eth_count + dep + contr.lin , a.design)
olr3_5 = svyolr(bmi_t ~ stair + age + eth_count + dep + contr.lin , a.design)
olr3_6 = svyolr(bmi_t ~ active + age + eth_count + dep + contr.lin , a.design)
olr3_7 = svyolr(bmi_t ~ native + age + eth_count + dep + contr.lin , a.design)
olr3_8 = svyolr(bmi_t ~ UA + age + eth_count + dep + contr.lin , a.design)
olr3_9 = svyolr(bmi_t ~ hhinc + age + eth_count + dep + contr.lin , a.design)
olr3_10 = svyolr(bmi_t ~ fruit + age + eth_count + dep + contr.lin , a.design)

summy(olr3_1)
```

```
summy(olr3_2)
summy(olr3_3)
summy(olr3_4)
summy(olr3_5)
summy(olr3_6)
summy(olr3_7)
summy(olr3_8)
summy(olr3_9)
summy(olr3_10)
# Difficulty climbing several steps of stairs had the strongest effect


## 4th step
olr4_1 = svyolr(bmi_t ~ gender + stair + age + eth_count + dep + contr.lin, a.design)
olr4_2 = svyolr(bmi_t ~ edu + stair + age + eth_count + dep + contr.lin, a.design)
olr4_3 = svyolr(bmi_t ~ sedentary + stair + age + eth_count + dep + contr.lin, a.design)
olr4_4 = svyolr(bmi_t ~ smoke + stair + age + eth_count + dep + contr.lin, a.design)
olr4_5 = svyolr(bmi_t ~ active + stair + age + eth_count + dep + contr.lin, a.design)
olr4_6 = svyolr(bmi_t ~ native + stair + age + eth_count + dep + contr.lin, a.design)
olr4_7 = svyolr(bmi_t ~ UA + stair + age + eth_count + dep + contr.lin, a.design)
olr4_8 = svyolr(bmi_t ~ hhinc + stair + age + eth_count + dep + contr.lin, a.design)

summy(olr4_1)
summy(olr4_2)
summy(olr4_3)
summy(olr4_4)
summy(olr4_5)
summy(olr4_6)
summy(olr4_7)
summy(olr4_8)

# Sedentary lifestyle lost its significance
# Gender had the strongest effect


## 5th step
olr5_1 = svyolr(bmi_t ~ edu + gender + stair + age + eth_count + dep + contr.lin, a.design)
olr5_2 = svyolr(bmi_t ~ smoke + gender + stair + age + eth_count + dep + contr.lin, a.design)
olr5_3 = svyolr(bmi_t ~ active + gender + stair + age + eth_count + dep + contr.lin, a.design)
olr5_4 = svyolr(bmi_t ~ native + gender + stair + age + eth_count + dep + contr.lin, a.design)
```

```
olr5_5 = svyolr(bmi_t ~ UA + gender + stair + age + eth_count + dep + contr.lin , a.design)
olr5_6 = svyolr(bmi_t ~ hhinc + gender + stair + age + eth_count + dep + contr.lin , a.design)

summy(olr5_1)
summy(olr5_2)
summy(olr5_3)
summy(olr5_4)
summy(olr5_5)
summy(olr5_6)
# Ex-smoker had the strongest effect


## 6th step
olr6_1 = svyolr(bmi_t ~ edu + smoke + gender + stair + age + eth_count + dep + contr.lin , a.design)
olr6_2 = svyolr(bmi_t ~ active + smoke + gender + stair + age + eth_count + dep + contr.lin , a.
    design)
olr6_3 = svyolr(bmi_t ~ native + smoke + gender + stair + age + eth_count + dep + contr.lin , a.
    design)
olr6_4 = svyolr(bmi_t ~ UA + smoke + gender + stair + age + eth_count + dep + contr.lin , a.design)
olr6_5 = svyolr(bmi_t ~ hhinc + smoke + gender + stair + age + eth_count + dep + contr.lin , a.design
    )

summy(olr6_1)
summy(olr6_2)
summy(olr6_3)
summy(olr6_4)
summy(olr6_5)
# Urban/rural area lose its significance
# Household income had the strongest effect


## 7th step
olr7_1 = svyolr(bmi_t ~ edu + hhinc + smoke + gender + stair + age + eth_count + dep + contr.lin , a.
    design)
olr7_2 = svyolr(bmi_t ~ native + hhinc + smoke + gender + stair + age + eth_count + dep + contr.lin ,
     a.design)
olr7_3 = svyolr(bmi_t ~ active + hhinc + smoke + gender + stair + age + eth_count + dep + contr.lin ,
     a.design)

summy(olr7_1)
```

```r
summy(olr7_2)
summy(olr7_3)
# Education had the strongest effect


## 8th step
olr8_1 = svyolr(bmi_t ~ active + edu + hhinc + smoke + gender +
                   stair + age + eth_count + dep + contr.lin , a.design)
olr8_2 = svyolr(bmi_t ~ native + edu + hhinc + smoke + gender +
                   stair + age + eth_count + dep + contr.lin , a.design)

summy(olr8_1)
summy(olr8_2)
# Physical activity had the strongest effect


## 9th step
olr9 = svyolr(bmi_t ~ native + active + edu + hhinc + smoke + gender +
                   stair + age + eth_count + dep + contr.lin , a.design)

summy(olr9 , round=4)
# Migration status remained significant
# olr9 is the final model

# calculating one-tailed z-test for the deprivation quintiles
# Alt. Hypothesis = The likelihood of being in the higher BMI tertiles is higher for people living
    in areas with higher depriation
pnorm(coefficients(summary(olr9))[19,3], lower.tail=FALSE) # quintile 2
pnorm(coefficients(summary(olr9))[20,3], lower.tail=FALSE) # quintile 3
pnorm(coefficients(summary(olr9))[21,3], lower.tail=FALSE) # quintile 4
pnorm(coefficients(summary(olr9))[22,3], lower.tail=FALSE) # quintile 5



#### Proportional Odds Assumption - Adult ####

# Binomial regression using contrast
# ex1 assessing the odds of being in the Mid or Top tertile vs. the Bottom tertile
# ex2 assessing the odds of being in the Top tertile vs. Bottom or Mid tertile
ex1 <- svyglm(I(bmi_t>"Bottom Tertile") ~ dep + contr.lin + eth_count + age + stair +
```

```r
                 gender + smoke + hhinc + edu + active + native , a.design , family = "binomial")
ex2 <- svyglm(I(bmi_t>"Mid Tertile") ~ dep + contr.lin + eth_count + age + stair +
                 gender + smoke + hhinc + edu + active + native , a.design , family = "binomial")

# plot them
plot(coef(ex1)[-1], xlim=c(0,30), ylim=c(-1,2), type="l", ylab="Beta", main="Comparison of Betas for
    \nProportional Odds Assumption")
lines(coef(ex2)[-1], col=2)

# Standard Errors and Coefficients
rbind(SE(ex1),SE(ex2))[,-1]
rbind(coef(ex1),coef(ex2))[,-1]

# odds ratios
rbind(exp(coef(ex1)),exp(coef(ex2)))[,-1]

# 2.5% CL odds ratio
rbind(exp(coef(ex1)-SE(ex1)), exp(coef(ex2)-SE(ex2)))[,-1]

# 97.5% CL odds ratio
rbind(exp(coef(ex1)+SE(ex1)), exp(coef(ex2)+SE(ex2)))[,-1]


#### Bivariate Analysis - Child ####
# All possible combination of explanatory variables with the outcome variable
# ANOVA = numeric x categorical; chi-squared = categorical x categorical
# Excluding missing BMI value

## Unweigthed Stacked plots
# Age X bmic (bmi category)
cage = box_notitle(child , "bmic", "age", "BMI Category", "Age", min=2, max = 18, bins = 2, gap =
    1.5)
summary(lm(age ~ bmic, data = child))
anova_table(child$age, child$bmic)

t = aggregate(age ~ bmic, child , function(x) c(mean=mean(x),sd=sd(x)))

# Sex
csex = stacked_notitle(child , "bmic", "gender", "Sex", "BMI Category")
t = table(child$bmic, child$gender)
```

```r
t = prop.table(table(child$bmic, child$gender), 1)

# Ethnicity
ceth = stacked_notitle(child, "bmic", "eth_count", "Ethnicity", "BMI Category")
t = table(child$bmic, child$eth_count)
t = prop.table(table(child$bmic, child$eth_count), 1)

# Deprivation Quintile
cdep = stacked_notitle(child, "bmic", "dep", "Deprivation Quintile", "BMI Category")
t = table(child$bmic, child$dep)
t = prop.table(table(child$bmic, child$dep), 1)

# Urban/Rural
cua = stacked_notitle(child, "bmic", "UA", "Urban/Rural Area", "BMI Category")
t = table(child$bmic, child$UA)
t = prop.table(table(child$bmic, child$UA), 1)

# Household Income
cinc = stacked_notitle(child, "bmic", "hhinc", "Household Income", "BMI Category")
t = table(child$bmic, child$hhinc)
t = prop.table(table(child$bmic, child$hhinc), 1)

# Education
cedu = stacked_notitle(child, "bmic", "edu", "Education", "BMI Category")
t = table(child$bmic, child$edu)
t = prop.table(table(child$bmic, child$edu), 1)

# Fruit Guideline
cfruit = stacked_notitle(child, "bmic", "fruit", "Fruit Guideline", "BMI Category")
t = table(child$bmic, child$fruit)
t = prop.table(table(child$bmic, child$fruit), 1)

# Vegetable Guideline
cvege = stacked_notitle(child, "bmic", "veges", "Vegetable Guideline", "BMI Category")
t = table(child$bmic, child$veges)
t = prop.table(table(child$bmic, child$veges), 1)

# Soft Drink Consumption
csoft = stacked_notitle(child, "bmic", "softd", "Soft Drink Consumption", "BMI Category")
t = table(child$bmic, child$softd)
```

```r
t = prop.table(table(child$bmic, child$softd), 1)

# Fast Food Consumption
cfast = stacked_notitle(child, "bmic", "fastf", "Fast Food Consumption", "BMI Category")
t = table(child$bmic, child$fastf)
t = prop.table(table(child$bmic, child$fastf), 1)


# arrange the charts
grid.arrange(cage, csex, ceth, cdep, cua, cinc, cedu, cfruit, cvege, csoft, cfast,
             ncol = 3, nrow=4, padding=unit(3, "line"))



## Chi-Squared tests (unweighted)
label = c("Sex", "Ethnicity", "Deprivation Quintile", "Urban/Rural Area", "Household Income","
    Education","Fruit Guideline",
          "Vegetable Guideline","Soft Drink Consumption","Fast Food Consmption")

table = rbind(chix_table(child$bmic, child$gender),chix_table(child$bmic, child$eth_count),chix_
    table(child$bmic, child$dep),
      chix_table(child$bmic, child$UA),chix_table(child$bmic, child$hhinc),chix_table(child$bmic,
          child$edu),
      chix_table(child$bmic, child$fruit),chix_table(child$bmic, child$veges),chix_table(child$bmic,
           child$softd),
      chix_table(child$bmic, child$fastf))

biv_table_child = cbind(label, table)

# remove unused object from memory
rm(label)
rm(table)


#### Relevel and set reference level - Child ####

child$veges = droplevels(child$veges)
child$eth_count = relevel(child$eth_count, ref="European Only")
child$hhinc = relevel(child$hhinc, ref=">$70,000")
child$bmic = ordered(child$bmic)
```

```r
child$year = as.factor(child$year)
child$year = relevel(child$year, ref="2014/15")
child$softd = relevel(child$softd, ref="0/week")
child$fastf = relevel(child$fastf, ref="0/week")

child$soft_order = factor(child$softd, levels=c("0/week", "1/week", "2-3/week", "4+/week"), ordered=
    TRUE)
child$fast_order = factor(child$fastf, levels=c("0/week", "1/week", "2-3/week", "4+/week"), ordered=
    TRUE)


#### Survey Design Object - Child ####

# missing sample weights
child$finalwgt = as.numeric(child$finalwgt)
sum(is.na(child$finalwgt)) # 4 NA

# remove them from analysis
child = subset(child, !is.na(finalwgt))

# creating a linear contrast
contr.lin <- contr.poly(5)[,1]

# bind the contrast to the data set
child$contr.lin <- contr.lin[ match (child$year, c("2006/07", "2011/12",
                                                   "2012/13","2013/14", "2014/15"))]

rm(contr.lin)

# complex survey design
c.design <- svydesign(
  id = ~cluster,
  strata = ~interaction(strata,year),
  data = child,
  weights = ~finalwgt,
  nest = TRUE
)

summary(c.design)

```

```
#### Obesity Prevalence by Year – Child ####


## BMI categories by year (non-age standardised, using survey weights)

bmic_year <- svyby(~bmic, ~year, svymean, design= c.design)
bmic_year <- gather(data.frame(bmic_year), bmic, prop, 2:4, factor_key=TRUE)

bmic_year = bmic_year %>%
  group_by(year) %>%
  mutate(pos = cumsum(prop) - 0.5*prop) # calculating position

bmic_year$label_p = paste0(sprintf("%.1f", bmic_year$prop*100),"%")

# bar plot
bmi_child_year = ggplot(bmic_year, aes(x=year, y=prop, fill=bmic, levels=bmic)) +
  geom_bar(stat="identity", width = 0.8) +
  labs(title= "Child Data", x= "Year", y= "Density") +
  scale_fill_discrete(name = "BMI Category", labels = c("Average","Overweight","Obese")) +
  theme(axis.title = element_text(size = 12, face="bold"),
        plot.title = element_text(size = 14, face="bold", hjust=0.5),
        legend.title = element_text(size = 12, face="bold"),
        axis.text.x = element_text(size = 11),
        axis.text.y = element_text(size = 11)) +
  geom_text(aes(label=label_p), position = position_stack(vjust = 0.5), size=4, colour = "white") +
  scale_x_discrete(limits=c("2006/07","2011/12","2012/13","2013/14","2014/15"),
                   labels=c("'06/07","'11/12","'12/13","'13/14", "'14/15"))


plot_grid(bmic_adult_year, bmi_child_year, align = "h", rel_widths=c(0.8,1))


## calculate bmi categories by ethnicity and year ()
# create age group variable
child$agegrp = cut(child$age, breaks = c(seq(0,85, by = 5), Inf), right = FALSE)


# calculate rates per 100 population (or proportion of BMI categories by year, ethnicity, age group)
obes_c = child %>%
```

```
  group_by(year, agegrp, eth_count, bmic) %>%
  filter(!is.na(eth_count)) %>%
  summarise(n=n()) %>%
  mutate(prop = n/sum(n))


# add scandinavian population proportion and rescale them so that they add up to 100%
scand_prop_child = c(0.08,0.07,0.07,0.07)
scand_prop_child = scand_prop_child*(1/(0.08 + 0.07 + 0.07 + 0.07))

# match them up
obes_c$scand_prop_child = scand_prop_child[ match (obes_c$agegrp, c("[0,5)", "[5,10)",  "[10,15)",
    "[15,20)"))]


# calculate age-specific rate
obes_c$age_r = obes_c$prop*obes_c$scand_prop_child


cage_std_plot = aggregate(age_r ~ year+eth_count+bmic, FUN=sum, data=obes_c, na.rm=TRUE)

cage_std_plot = spread(cage_std_plot, bmic, age_r)

cage_std_plot[,3:5] = sapply(cage_std_plot[,3:5], function(x) x*100)

cage_std_plot$over_obes = cage_std_plot$Overweight+cage_std_plot$Obese


## Obesity rate by year and ethncity
cobes_plot = ggplot(cage_std_plot, aes(x=year, y=Obese, shape=eth_count, group=eth_count), stat="
    identity") +
  geom_line(size=1) +
  geom_point(size=2) +
  theme_light() +
  scale_x_discrete(limits=c("2006/07","2011/12","2012/13","2013/14","2014/15")) +
  scale_y_continuous(limits=c(0,80),breaks=seq(0,100,10)) +
  labs(title="Child Data", x = "Year", y = "Percentage", shape="Ethnicity") +
  theme(axis.title = element_text(size = 14, face="bold"),
        legend.title = element_text(size = 14, face="bold"),
        legend.text = element_text(size = 12),
```

```
        axis.text.x = element_text(size = 12),
        axis.text.y = element_text(size = 12),
        plot.title = element_text(size = 16, face="bold", hjust=0.5))

plot_grid(aobes_plot, cobes_plot, align = "h", rel_widths=c(0.8,1))



## Overweight and obesity rate by year and ethnicity
covob_plot = ggplot(cage_std_plot, aes(x=year, y=Overweight+Obese, shape=eth_count, group=eth_count)
    , stat="identity") +
  geom_line(size=1) +
  geom_point(size=2) +
  theme_light() +
  scale_x_discrete(limits=c("2006/07","2011/12","2012/13","2013/14","2014/15")) +
  scale_y_continuous(limits=c(20,100),breaks=seq(0,100,10)) +
  labs(title="Child Data", x = "Year", y = "Percentage", shape="Ethnicity") +
  theme(axis.title = element_text(size = 14, face="bold"),
        legend.title = element_text(size = 14, face="bold"),
        legend.text = element_text(size = 12),
        axis.text.x = element_text(size = 12),
        axis.text.y = element_text(size = 12),
        plot.title = element_text(size = 16, face="bold", hjust=0.5))

plot_grid(aovob_plot, covob_plot, align = "h", rel_widths=c(0.8,1))



#### Descriptive Analysis across survey periods - Child ####

# total sample size on each year
child %>%
  group_by(year) %>%
  summarise(row = length(age))

# age summary
svyquantile(~age, subset(c.design, year == "2006/07"), c(0,.25,.5,.75,1), ci=TRUE)
svyquantile(~age, subset(c.design, year == "2011/12"), c(0,.25,.5,.75,1), ci=TRUE)
svyquantile(~age, subset(c.design, year == "2012/13"), c(0,.25,.5,.75,1), ci=TRUE)
svyquantile(~age, subset(c.design, year == "2013/14"), c(0,.25,.5,.75,1), ci=TRUE)
```

```r
svyquantile(~age, subset(c.design, year == "2014/15"), c(0,.25,.5,.75,1), ci=TRUE)


# categorical variables
svyby(~bmic, ~year, svymean, design=c.design)
svyby(~gender, ~year, svymean, design=c.design)
svyby(~eth_count, ~year, svymean, design=c.design)
svyby(~dep, ~year, ~year,svymean, design=c.design)
svyby(~UA, ~year, svymean, design=c.design)
svyby(~hhinc, ~year, svymean, design=c.design)
svyby(~edu, ~year, svymean, design=c.design)
svyby(~fruit, ~year, svymean, design=c.design)
svyby(~veges, ~year, svymean, design=c.design)
svyby(~softd, ~year, svymean, design=c.design)
svyby(~fastf, ~year, svymean, design=c.design)


#### Trends across survey periods - Child ####
# reference level = 2014/15
summy(svyglm(I(bmic == "Obese") ~ year, c.design))
c.design$variables$year = relevel(c.design$variables$year, ref= "2011/12")
summy(svyglm(I(bmic == "Obese") ~ year, c.design)) # no diff 2011/12 vs 2006/07
c.design$variables$year = relevel(c.design$variables$year, ref= "2012/13")
summy(svyglm(I(bmic == "Obese") ~ year, c.design)) # sig. diff: 2012/13 vs 2006/07

# reset the reference group to "2014/15"
c.design$variables$year = factor(c.design$variables$year, levels=c("2002/03",
                                "2006/07", "2011/12", "2012/13","2013/14", "2014/15"))
c.design$variables$year = relevel(c.design$variables$year, ref= "2014/15")


summy(svyglm(I(eth_count == "European Only") ~ year, c.design))
summy(svyglm(I(eth_count == "Other") ~ year, c.design))
summy(svyglm(I(UA == "Urban") ~ year, c.design))
summy(svyglm(I(edu == "Tertiary") ~ year, c.design))
summy(svyglm(I(soft_order > "1/week") ~ year, c.design))
summy(svyglm(I(fast_order > "1/week") ~ year, c.design))


#### Forward Step Proportional Odds Logistic Regression - Child ####
```

```
# Add one covariate at a time, add the strongest covariate on each step, and remove any non
    significant variable (p < 0.1)

## 1st step
# add deprivation quintile and linear contrasts, variables of interest
colr1_1 = svyolr(bmic ~ dep + contr.lin, c.design)

summy(colr1_1)


## 2nd step
colr2_0 = svyolr(bmic ~ gender + dep + contr.lin, c.design)
colr2_1 = svyolr(bmic ~ age + dep + contr.lin, c.design)
colr2_2 = svyolr(bmic ~ eth_count + dep + contr.lin, c.design)
colr2_3 = svyolr(bmic ~ hhinc + dep + contr.lin, c.design)
colr2_4 = svyolr(bmic ~ edu + dep + contr.lin, c.design)
colr2_5 = svyolr(bmic ~ fruit + dep + contr.lin, c.design)
colr2_6 = svyolr(bmic ~ veges + dep + contr.lin, c.design)
colr2_7 = svyolr(bmic ~ softd + dep + contr.lin, c.design)
colr2_8 = svyolr(bmic ~ fastf + dep + contr.lin, c.design)
colr2_9 = svyolr(bmic ~ UA + dep + contr.lin, c.design)

summy(colr2_0)
summy(colr2_1)
summy(colr2_2)
summy(colr2_3)
summy(colr2_4)
summy(colr2_5)
summy(colr2_6)
summy(colr2_7)
summy(colr2_8)
summy(colr2_9)
# fruit, veges and urban/rural area were not significant
# Ethnicity had the strongest effect


## 3rd step
colr3_1 = svyolr(bmic ~ age + eth_count + dep + contr.lin, c.design)
colr3_2 = svyolr(bmic ~ hhinc + eth_count + dep + contr.lin, c.design)
colr3_3 = svyolr(bmic ~ edu + eth_count + dep + contr.lin, c.design)
```

```r
colr3_4 = svyolr(bmic ~ softd + eth_count + dep + contr.lin , c.design)
colr3_5 = svyolr(bmic ~ fastf + eth_count + dep + contr.lin , c.design)
colr3_6 = svyolr(bmic ~ gender + eth_count + dep + contr.lin , c.design)

summy(colr3_1)
summy(colr3_2)
summy(colr3_3)
summy(colr3_4)
summy(colr3_5)
summy(colr3_6)
# household income lost became insignificant
# Age had the strongest effect


## 4th step
colr4_1 = svyolr(bmic ~ edu + age + eth_count + dep + contr.lin , c.design)
colr4_2 = svyolr(bmic ~ softd + age + eth_count + dep + contr.lin , c.design)
colr4_3 = svyolr(bmic ~ fastf + age + eth_count + dep + contr.lin , c.design)
colr4_4 = svyolr(bmic ~ gender + age + eth_count + dep + contr.lin , c.design)

summy(colr4_4)
summy(colr4_1)
summy(colr4_2)
summy(colr4_3)
# Education had the strongest effect


## 5th step
colr5_1 = svyolr(bmic ~ softd + edu + age + eth_count + dep + contr.lin , c.design)
colr5_2 = svyolr(bmic ~ fastf + edu + age + eth_count + dep + contr.lin , c.design)
colr5_3 = svyolr(bmic ~ gender + edu + age + eth_count + dep + contr.lin , c.design)

summy(colr5_1)
summy(colr5_2)
summy(colr5_3)
# soft drink consumption had the strongest effect


## 6th step
colr6_1 = svyolr(bmic ~ fastf + softd + edu + age + eth_count + dep + contr.lin , c.design)
```

```
colr6_2 = svyolr(bmic ~ gender + softd + edu + age + eth_count + dep + contr.lin , c.design)

summy(colr6_1)
summy(colr6_2)
# fast food consumption lost was not significant , maybe due to colinearity with soft drink
    consumption


## Model Fit Tests
## Model 1 (0.05 rule): bmic ~ softd + edu + age + eth_count + dep + contr.lin , c.design
## Model 2 (0.1 rule): bmic ~ gender + softd + edu + age + eth_count + dep + contr.lin , c.design
## Model 3 (0.2 rule): bmic ~ fastf + gender + softd + edu + age + eth_count + dep + contr.lin , c.
    design

# does adding gender into model 1 is better?
model1 = svyolr(bmic ~ softd + edu + age + eth_count + dep + contr.lin , c.design)
# Wald test
regTermTest(colr6_2, ~gender, df=NULL) # yes

summary(model1)
summary(colr6_2) # model 2
# minimal difference in the beta and SE of both model
# SE in model 2 tends to be slightly higher than model 1


# does adding fast food into model 2 is better?
colr7_1 = svyolr(bmic ~ fastf + gender + softd + edu + age + eth_count + dep + contr.lin , c.design)
regTermTest(colr7_1, ~fastf , df=NULL) # no

# col6_2 is the final model
# one-tailed z-test for deprivation quintiles
pnorm(coefficients(summary(colr6_2))[13,3], lower.tail=FALSE) # quintile 2
pnorm(coefficients(summary(colr6_2))[14,3], lower.tail=FALSE) # quintile 3
pnorm(coefficients(summary(colr6_2))[15,3], lower.tail=FALSE) # quintile 4
pnorm(coefficients(summary(colr6_2))[16,3], lower.tail=FALSE) # quintile 5


#### Proportional Odds Assumption - Child ####

## Binomial regression using contrast
```

```
# cex1 is asessing the odds of being in the overweight/obese vs. average category
# cex2 is assessing the odds of being in the obese vs. average/overweight category
cex1 = svyglm(I(bmic>"Average") ~ dep + contr.lin + eth_count + age + edu + softd + gender, c.design
    , family = "binomial")
cex2 = svyglm(I(bmic>"Overweight") ~ dep + contr.lin + eth_count + age + edu + softd + gender, c.
    design, family="binomial")

plot(coef(cex1)[-1], xlim=c(0,20), ylim=c(-1,2), type="l", ylab="Beta", main="Comparison of Betas
    for \nProportional Odds Assumption")
lines(coef(cex2)[-1], col=2)

## Beta coefficients and standard errors
rbind(coef(cex1),coef(cex2))[,-1]
rbind(SE(cex1),SE(cex2))[,-1]
# higher beta of being in the obese vs average/overweight category compared with being in overweight
    /obese vs average for the following variables: soft drink 4+/week, all ethnicities, and
    deprivation.

## odds ratios
rbind(exp(coef(cex1)),exp(coef(cex2)))[,-1]

# 2.5% CL for OR
rbind(exp(coef(cex1)-SE(cex1)), exp(coef(cex2)-SE(cex2)))[,-1]

# 97.5% CL for OR
rbind(exp(coef(cex1)+SE(cex1)), exp(coef(cex2)+SE(cex2)))[,-1]
```

Listing A.3: Bivariate Analyses - BMI

```r

#### BIVARIATE ANALYSES
## This section contains unweighted plots and appropriate statistical tests on all of the
    combination between the explanatory variables and the outcome variables


## Merge
ac06 <- merge(a06, c06, all=TRUE)
ac11 <- merge(a11, c11, all=TRUE)
ac12 <- merge(a12, c12, all=TRUE)
ac13 <- merge(a13, c13, all=TRUE)
ac14 <- merge(a14, c14, all=TRUE)

# add year variable to all data sets
hs02$year <- rep("2002/03")
ac06$year <- rep("2006/07")
ac11$year <- rep("2011/12")
ac12$year <- rep("2012/13")
ac13$year <- rep("2013/14")
ac14$year <- rep("2014/15")

# Split them into child data (2-17 years old) and adult data (18+ years old)
adult <- subset(hs02, age >=18) %>%
         merge(subset(ac06, age >=18), all=TRUE) %>%
         merge(subset(ac11, age >=18), all=TRUE) %>%
         merge(subset(ac12, age >=18), all=TRUE) %>%
         merge(subset(ac13, age >=18), all=TRUE) %>%
         merge(subset(ac14, age >=18), all=TRUE)

child <- subset(ac06, age >=2 & age <18) %>%
         merge(subset(ac11, age >=2 & age <18), all=TRUE) %>%
         merge(subset(ac12, age >=2 & age <18), all=TRUE) %>%
         merge(subset(ac13, age >=2 & age <18), all=TRUE) %>%
         merge(subset(ac14, age >=2 & age <18), all=TRUE)


# Exclude missing BMI values in the Adult data
adult <- subset(adult, !is.na(bmiscale))

```

```r

# Transform them into tertiles
adult$bmi_t <- factor(ntile(adult$bmiscale, 3),
                      labels=c("Bottom Tertile","Mid Tertile","Top Tertile"), ordered=TRUE)


# Include only relevant variables
adult <- subset(adult, select = c("age","bmic","gender","bmiscale", "UA", "bmi_t",
                                  "dep", "hhinc","edu", "active", "native",
                                  "sedentary","stair","fruit","veges",
                                  "year","smoke", "eth_count", "finalwgt", "cluster", "strata",
                                  "haz_drinker_all"))


# also exclude missing BMI value in the Child data and include relevant variables
child <- subset(child, !is.na(bmic))

child <- subset(child, select=c("age","bmic","gender","bmiscale", "UA",
                                  "dep", "hhinc","edu","fruit","veges",
                                  "year","softd","fastf", "eth_count",
                                  "finalwgt", "strata", "cluster"))



## Adult Data
## Subset the data based on year
adult02 = subset(adult, year == "2002/03")
adult06 = subset(adult, year == "2006/07")
adult11 = subset(adult, year == "2011/12")
adult12 = subset(adult, year == "2012/13")
adult13 = subset(adult, year == "2013/14")
adult14 = subset(adult, year == "2014/15")


# AGE
grid.arrange(box_plot_median(adult02, "bmi_t", "age", "Age", "BMI Tertile Group", "2002/03"),
             box_plot_median(adult06, "bmi_t", "age", "Age", "BMI Tertile Group", "2006/07"),
             box_plot_median(adult11, "bmi_t", "age", "Age", "BMI Tertile Group", "2011/12"),
             box_plot_median(adult12, "bmi_t", "age", "Age", "BMI Tertile Group", "2012/13"),
             box_plot_median(adult13, "bmi_t", "age", "Age", "BMI Tertile Group", "2013/14"),
```

```
            box_plot_median(adult14, "bmi_t", "age", "Age", "BMI Tertile Group", "2014/15"),
            ncol=3, nrow=2, padding = unit(3, "line"))


rbind(anova_table(adult02$age, adult02$bmi_t),
anova_table(adult06$age, adult06$bmi_t),
anova_table(adult11$age, adult11$bmi_t),
anova_table(adult12$age, adult12$bmi_t),
anova_table(adult13$age, adult13$bmi_t),
anova_table(adult14$age, adult14$bmi_t))


# SEX
xt_2x3(adult02$gender, adult02$bmi_t)
xt_2x3(adult06$gender, adult06$bmi_t)
xt_2x3(adult11$gender, adult11$bmi_t)
xt_2x3(adult12$gender, adult12$bmi_t)
xt_2x3(adult13$gender, adult13$bmi_t)
xt_2x3(adult14$gender, adult14$bmi_t)

rbind(chix_table(adult02$gender, adult02$bmi_t),
      chix_table(adult06$gender, adult06$bmi_t),
      chix_table(adult11$gender, adult11$bmi_t),
      chix_table(adult12$gender, adult12$bmi_t),
      chix_table(adult13$gender, adult13$bmi_t),
      chix_table(adult14$gender, adult14$bmi_t))

grid.arrange(stacked_plot(adult02, "gender", "bmi_t", "Sex Group", "BMI Tertile", "2002/03", ">=18 yo"),
            stacked_plot(adult06, "gender", "bmi_t", "Sex Group", "BMI Tertile", "2006/07", ">=18 yo"),
            stacked_plot(adult11, "gender", "bmi_t", "Sex Group", "BMI Tertile", "2011/12", ">=18 yo"),
            stacked_plot(adult12, "gender", "bmi_t", "Sex Group", "BMI Tertile", "2012/13", ">=18 yo"),
            stacked_plot(adult13, "gender", "bmi_t", "Sex Group", "BMI Tertile", "2013/14", ">=18 yo"),
            stacked_plot(adult14, "gender", "bmi_t", "Sex Group", "BMI Tertile", "2014/15", ">=18 yo"),
            ncol=3, nrow=2, padding= unit(3, "line"))
```

```
# ETHNICITY
table(adult02$eth_count, adult02$bmi_t)
table(adult06$eth_count, adult06$bmi_t)
table(adult11$eth_count, adult11$bmi_t)
table(adult12$eth_count, adult12$bmi_t)
table(adult13$eth_count, adult13$bmi_t)
table(adult13$eth_count, adult13$bmi_t)

rbind(chix_table(adult02$eth_count, adult02$bmi_t),
      chix_table(adult06$eth_count, adult06$bmi_t),
      chix_table(adult11$eth_count, adult11$bmi_t),
      chix_table(adult12$eth_count, adult12$bmi_t),
      chix_table(adult13$eth_count, adult13$bmi_t),
      chix_table(adult14$eth_count, adult14$bmi_t))

grid.arrange(stacked_plot(adult02, "eth_count", "bmi_t", "Ethnic Group", "BMI Tertile", "2002/03", "
    >=18 yo"),
             stacked_plot(adult06, "eth_count", "bmi_t", "Ethnic Group", "BMI Tertile", "2006/07", "
                 >=18 yo"),
             stacked_plot(adult11, "eth_count", "bmi_t", "Ethnic Group", "BMI Tertile", "2011/12", "
                 >=18 yo"),
             stacked_plot(adult12, "eth_count", "bmi_t", "Ethnic Group", "BMI Tertile", "2012/13", "
                 >=18 yo"),
             stacked_plot(adult13, "eth_count", "bmi_t", "Ethnic Group", "BMI Tertile", "2013/14", "
                 >=18 yo"),
             stacked_plot(adult14, "eth_count", "bmi_t", "Ethnic Group", "BMI Tertile", "2014/15", "
                 >=18 yo"),
             ncol=3, nrow=2, padding=unit(3,"line"))


# DEPRIVATION
xt_5x3(adult02$Nzdep01_quintiles, adult02$bmi_t)
xt_5x3(adult06$nzdep06_quintile, adult06$bmi_t)
xt_5x3(adult11$nzdep06_quintile, adult11$bmi_t)
xt_5x3(adult12$nzdep06_quintile, adult12$bmi_t)
xt_5x3(adult13$nzdep13_quintile, adult13$bmi_t)
xt_5x3(adult14$nzdep13_quintile, adult14$bmi_t)

```

```r
rbind(chix_table(adult02$Nzdep01_quintiles, adult02$bmi_t),
      chix_table(adult06$nzdep06_quintile, adult06$bmi_t),
      chix_table(adult11$nzdep06_quintile, adult11$bmi_t),
      chix_table(adult12$nzdep06_quintile, adult12$bmi_t),
      chix_table(adult13$nzdep13_quintile, adult13$bmi_t),
      chix_table(adult14$nzdep13_quintile, adult14$bmi_t))

grid.arrange(stacked_plot(adult02, "Nzdep01_quintiles", "bmi_t", "Deprivation Quintile", "BMI
    Tertile", "2002/03"),
            stacked_plot(adult06, "nzdep06_quintile", "bmi_t", "Deprivation Quintile", "BMI Tertile"
                , "2006/07"),
            stacked_plot(adult11, "nzdep06_quintile", "bmi_t", "Deprivation Quintile", "BMI Tertile"
                , "2011/12"),
            stacked_plot(adult12, "nzdep06_quintile", "bmi_t", "Deprivation Quintile", "BMI Tertile"
                , "2012/13"),
            stacked_plot(adult13, "nzdep13_quintile", "bmi_t", "Deprivation Quintile", "BMI Tertile"
                , "2013/14"),
            stacked_plot(adult14, "nzdep13_quintile", "bmi_t", "Deprivation Quintile", "BMI Tertile"
                , "2014/15"),
            ncol= 3, nrow=2, padding=unit(3,"line"))


# URBAN/RURAL
xt_2x3(adult02$UA, adult02$bmi_t)
xt_2x3(adult06$UA, adult06$bmi_t)
xt_2x3(adult14$UA, adult14$bmi_t)

rbind(chix_table(adult02$UA, adult02$bmi_t),
      chix_table(adult06$UA, adult06$bmi_t),
      chix_table(adult14$UA, adult14$bmi_t))

grid.arrange(stacked_plot(adult02, "UA", "bmi_t", "Urban/Rural Area", "BMI Tertile", "2002/03"),
             stacked_plot(adult06, "UA", "bmi_t", "Urban/Rural Area", "BMI Tertile", "2006/07"),
             stacked_plot(adult14, "UA", "bmi_t", "Urban/Rural Area", "BMI Tertile", "2014/15"),
             ncol=3, padding=unit(3,"line"))


# HOUSEHOLD INCOME
xt_4x3(adult02$hhinc, adult02$bmi_t)
```

```
xt_4x3(adult06$hhinc, adult06$bmi_t)
xt_4x3(adult11$hhinc, adult11$bmi_t)
xt_4x3(adult12$hhinc, adult12$bmi_t)
xt_4x3(adult13$hhinc, adult13$bmi_t)
xt_4x3(adult14$hhinc, adult14$bmi_t)

rbind(chix_table(adult02$hhinc, adult02$bmi_t),
      chix_table(adult06$hhinc, adult06$bmi_t),
      chix_table(adult11$hhinc, adult11$bmi_t),
      chix_table(adult12$hhinc, adult12$bmi_t),
      chix_table(adult13$hhinc, adult13$bmi_t),
      chix_table(adult14$hhinc, adult14$bmi_t))

grid.arrange(stacked_plot(adult02, "hhinc", "bmi_t", "Deprivation Quintile", "BMI Tertile", "2002/03"),
            stacked_plot(adult06, "hhinc", "bmi_t", "Deprivation Quintile", "BMI Tertile", "2006/07"),
            stacked_plot(adult11, "hhinc", "bmi_t", "Deprivation Quintile", "BMI Tertile", "2011/12"),
            stacked_plot(adult12, "hhinc", "bmi_t", "Deprivation Quintile", "BMI Tertile", "2012/13"),
            stacked_plot(adult13, "hhinc", "bmi_t", "Deprivation Quintile", "BMI Tertile", "2013/14"),
            stacked_plot(adult14, "hhinc", "bmi_t", "Deprivation Quintile", "BMI Tertile", "2014/15"),
            ncol= 3, nrow=2, padding=unit(3,"line"))



# EDUCATION
xt_3x3(adult02$edu, adult02$bmi_t)
xt_3x3(adult06$edu, adult06$bmi_t)
xt_3x3(adult11$edu, adult11$bmi_t)
xt_3x3(adult12$edu, adult12$bmi_t)
xt_3x3(adult13$edu, adult13$bmi_t)
xt_3x3(adult14$edu, adult14$bmi_t)

rbind(chix_table(adult02$edu, adult02$bmi_t),
      chix_table(adult06$edu, adult06$bmi_t),
      chix_table(adult11$edu, adult11$bmi_t),
```

```r
      chix_table(adult12$edu, adult12$bmi_t),
      chix_table(adult13$edu, adult13$bmi_t),
      chix_table(adult14$edu, adult14$bmi_t))

grid.arrange(stacked_plot(adult02, "edu", "bmi_t", "Educational Qualification", "BMI Tertile", "2002
    /03"),
            stacked_plot(adult06, "edu", "bmi_t", "Educational Qualification", "BMI Tertile", "2006/
                07"),
            stacked_plot(adult11, "edu", "bmi_t", "Educational Qualification", "BMI Tertile", "2011/
                12"),
            stacked_plot(adult12, "edu", "bmi_t", "Educational Qualification", "BMI Tertile", "2012/
                13"),
            stacked_plot(adult13, "edu", "bmi_t", "Educational Qualification", "BMI Tertile", "2013/
                14"),
            stacked_plot(adult14, "edu", "bmi_t", "Educational Qualification", "BMI Tertile", "2014/
                15"),
            ncol= 3, nrow=2, padding=unit(3,"line"))


# FRUIT GUIDELINE
xt_2x3(adult02$fruit, adult02$bmi_t)
xt_2x3(adult06$fruit, adult06$bmi_t)
xt_2x3(adult11$fruit, adult11$bmi_t)
xt_2x3(adult12$fruit, adult12$bmi_t)
xt_2x3(adult13$fruit, adult13$bmi_t)
xt_2x3(adult14$fruit, adult14$bmi_t)

rbind(chix_table(adult02$fruit, adult02$bmi_t),
      chix_table(adult06$fruit, adult06$bmi_t),
      chix_table(adult11$fruit, adult11$bmi_t),
      chix_table(adult12$fruit, adult12$bmi_t),
      chix_table(adult13$fruit, adult13$bmi_t),
      chix_table(adult14$fruit, adult14$bmi_t))

grid.arrange(stacked_plot(adult02, "fruit", "bmi_t", "Fruit Guideline", "BMI Tertile", "2002/03"),
            stacked_plot(adult06, "fruit", "bmi_t", "Fruit Guideline", "BMI Tertile", "2006/07"),
            stacked_plot(adult11, "fruit", "bmi_t", "Fruit Guideline", "BMI Tertile", "2011/12"),
            stacked_plot(adult12, "fruit", "bmi_t", "Fruit Guideline", "BMI Tertile", "2012/13"),
            stacked_plot(adult13, "fruit", "bmi_t", "Fruit Guideline", "BMI Tertile", "2013/14"),
            stacked_plot(adult14, "fruit", "bmi_t", "Fruit Guideline", "BMI Tertile", "2014/15"),
```

```
            ncol= 3, nrow=2, padding=unit(3,"line"))


# VEGETABLE GUIDELINE
xt_2x3(adult02$veges, adult02$bmi_t)
xt_2x3(adult06$veges, adult06$bmi_t)
xt_2x3(adult11$veges, adult11$bmi_t)
xt_2x3(adult12$veges, adult12$bmi_t)
xt_2x3(adult13$veges, adult13$bmi_t)
xt_2x3(adult14$veges, adult14$bmi_t)

rbind(chix_table(adult02$veges, adult02$bmi_t),
      chix_table(adult06$veges, adult06$bmi_t),
      chix_table(adult11$veges, adult11$bmi_t),
      chix_table(adult12$veges, adult12$bmi_t),
      chix_table(adult13$veges, adult13$bmi_t),
      chix_table(adult14$veges, adult14$bmi_t))

grid.arrange(stacked_plot(adult02, "veges", "bmi_t", "Vegetable Guideline", "BMI Tertile", "2002/03"
    ),
            stacked_plot(adult06, "veges", "bmi_t", "Vegetable Guideline", "BMI Tertile", "2006/07")
                ,
            stacked_plot(adult11, "veges", "bmi_t", "Vegetable Guideline", "BMI Tertile", "2011/12")
                ,
            stacked_plot(adult12, "veges", "bmi_t", "Vegetable Guideline", "BMI Tertile", "2012/13")
                ,
            stacked_plot(adult13, "veges", "bmi_t", "Vegetable Guideline", "BMI Tertile", "2013/14")
                ,
            stacked_plot(adult14, "veges", "bmi_t", "Vegetable Guideline", "BMI Tertile", "2014/15")
                ,
            ncol= 3, nrow=2, padding=unit(3,"line"))


# MIGRATION
xt_2x3(adult02$native, adult02$bmi_t)
xt_2x3(adult06$native, adult06$bmi_t)
xt_2x3(adult11$native, adult11$bmi_t)
xt_2x3(adult12$native, adult12$bmi_t)
test <- xt_2x3(adult13$native, adult13$bmi_t)
xt_2x3(adult14$native, adult14$bmi_t)
```

```
rbind(chix_table(adult02$native, adult02$bmi_t),
      chix_table(adult06$native, adult06$bmi_t),
      chix_table(adult11$native, adult11$bmi_t),
      chix_table(adult12$native, adult12$bmi_t),
      chix_table(adult13$native, adult13$bmi_t),
      chix_table(adult14$native, adult14$bmi_t))

grid.arrange(stacked_plot(adult02, "native", "bmi_t", "Migration Status", "BMI Tertile", "2002/03"),
             stacked_plot(adult06, "native", "bmi_t", "Migration Status", "BMI Tertile", "2006/07"),
             stacked_plot(adult11, "native", "bmi_t", "Migration Status", "BMI Tertile", "2011/12"),
             stacked_plot(adult12, "native", "bmi_t", "Migration Status", "BMI Tertile", "2012/13"),
             stacked_plot(adult13, "native", "bmi_t", "Migration Status", "BMI Tertile", "2013/14"),
             stacked_plot(adult14, "native", "bmi_t", "Migration Status", "BMI Tertile", "2014/15"),
             ncol= 3, nrow=2, padding=unit(3,"line"))


# DIFFICULTY CLIMBING STAIRS
xt_3x3(adult02$stair, adult02$bmi_t)
xt_3x3(adult06$stair, adult06$bmi_t)
xt_3x3(adult11$stair, adult11$bmi_t)
xt_3x3(adult12$stair, adult12$bmi_t)
xt_3x3(adult13$stair, adult13$bmi_t)
xt_3x3(adult14$stair, adult14$bmi_t)

rbind(chix_table(adult02$stair, adult02$bmi_t),
      chix_table(adult06$stair, adult06$bmi_t),
      chix_table(adult11$stair, adult11$bmi_t),
      chix_table(adult12$stair, adult12$bmi_t),
      chix_table(adult13$stair, adult13$bmi_t),
      chix_table(adult14$stair, adult14$bmi_t))

grid.arrange(stacked_plot(adult02, "stair", "bmi_t", "Difficulty Climbing Stairs", "BMI Tertile", "
    2002/03"),
             stacked_plot(adult06, "stair", "bmi_t", "Difficulty Climbing Stairs", "BMI Tertile", "
                2006/07"),
             stacked_plot(adult11, "stair", "bmi_t", "Difficulty Climbing Stairs", "BMI Tertile", "
                2011/12"),
             stacked_plot(adult12, "stair", "bmi_t", "Difficulty Climbing Stairs", "BMI Tertile", "
                2012/13"),
```

```r
            stacked_plot(adult13, "stair", "bmi_t", "Difficulty Climbing Stairs", "BMI Tertile", "2013/14"),
            stacked_plot(adult14, "stair", "bmi_t", "Difficulty Climbing Stairs", "BMI Tertile", "2014/15"),
            ncol= 3, nrow=2, padding=unit(3,"line"))


# SMOKING STATUS
xt_3x3(adult06$smoke, adult06$bmi_t)
xt_3x3(adult11$smoke, adult11$bmi_t)
xt_3x3(adult12$smoke, adult12$bmi_t)
xt_3x3(adult13$smoke, adult13$bmi_t)
xt_3x3(adult14$smoke, adult14$bmi_t)

rbind(chix_table(adult06$smoke, adult06$bmi_t),
      chix_table(adult11$smoke, adult11$bmi_t),
      chix_table(adult12$smoke, adult12$bmi_t),
      chix_table(adult13$smoke, adult13$bmi_t),
      chix_table(adult14$smoke, adult14$bmi_t))

grid.arrange(stacked_plot(adult06, "smoke", "bmi_t", "Smoking Status", "BMI Tertile", "2006/07"),
            stacked_plot(adult11, "smoke", "bmi_t", "Smoking Status", "BMI Tertile", "2011/12"),
            stacked_plot(adult12, "smoke", "bmi_t", "Smoking Status", "BMI Tertile", "2012/13"),
            stacked_plot(adult13, "smoke", "bmi_t", "Smoking Status", "BMI Tertile", "2013/14"),
            stacked_plot(adult14, "smoke", "bmi_t", "Smoking Status", "BMI Tertile", "2014/15"),
            ncol= 3, nrow=2, padding=unit(3,"line"))


# ALCOHOL PROBLEM
xt_2x3(adult02$haz_drinker_all, adult02$bmi_t)
xt_2x3(adult06$haz_drinker_all, adult06$bmi_t)
xt_2x3(adult11$haz_drinker_all, adult11$bmi_t)
xt_2x3(adult12$haz_drinker_all, adult12$bmi_t)
xt_2x3(adult13$haz_drinker_all, adult13$bmi_t)
xt_2x3(adult14$haz_drinker_all, adult14$bmi_t)

rbind(chix_table(adult02$haz_drinker_all, adult02$bmi_t),
      chix_table(adult06$haz_drinker_all, adult06$bmi_t),
      chix_table(adult11$haz_drinker_all, adult11$bmi_t),
      chix_table(adult12$haz_drinker_all, adult12$bmi_t),
```

```
      chix_table(adult13$haz_drinker_all, adult13$bmi_t),
      chix_table(adult14$haz_drinker_all, adult14$bmi_t))

grid.arrange(stacked_plot(adult02, "haz_drinker_all", "bmi_t", "Alcohol Problem", "BMI Tertile", "
    2002/03"),
            stacked_plot(adult06, "haz_drinker_all", "bmi_t", "Alcohol Problem", "BMI Tertile", "
                2006/07"),
            stacked_plot(adult11, "haz_drinker_all", "bmi_t", "Alcohol Problem", "BMI Tertile", "
                2011/12"),
            stacked_plot(adult12, "haz_drinker_all", "bmi_t", "Alcohol Problem", "BMI Tertile", "
                2012/13"),
            stacked_plot(adult13, "haz_drinker_all", "bmi_t", "Alcohol Problem", "BMI Tertile", "
                2013/14"),
            stacked_plot(adult14, "haz_drinker_all", "bmi_t", "Alcohol Problem", "BMI Tertile", "
                2014/15"),
            ncol= 3, nrow=2, padding=unit(3,"line"))


# PHYSICAL ACTIVITY
xt_2x3(adult02$active, adult02$bmi_t)
xt_2x3(adult06$active, adult06$bmi_t)
xt_2x3(adult11$active, adult11$bmi_t)
xt_2x3(adult12$active, adult12$bmi_t)
xt_2x3(adult13$active, adult13$bmi_t)
xt_2x3(adult14$active, adult14$bmi_t)

rbind(chix_table(adult02$active, adult02$bmi_t),
      chix_table(adult06$active, adult06$bmi_t),
      chix_table(adult11$active, adult11$bmi_t),
      chix_table(adult12$active, adult12$bmi_t),
      chix_table(adult13$active, adult13$bmi_t),
      chix_table(adult14$active, adult14$bmi_t))

grid.arrange(stacked_plot(adult02, "active", "bmi_t", "Physical Activity", "BMI Tertile", "2002/03")
    ,
            stacked_plot(adult06, "active", "bmi_t", "Physical Activity", "BMI Tertile", "2006/07"),
            stacked_plot(adult11, "active", "bmi_t", "Physical Activity", "BMI Tertile", "2011/12"),
            stacked_plot(adult12, "active", "bmi_t", "Physical Activity", "BMI Tertile", "2012/13"),
            stacked_plot(adult13, "active", "bmi_t", "Physical Activity", "BMI Tertile", "2013/14"),
            stacked_plot(adult14, "active", "bmi_t", "Physical Activity", "BMI Tertile", "2014/15"),
```

```
            ncol= 3, nrow=2, padding=unit(3,"line"))


# SEDENTARY LIFESTYLE
xt_2x3(adult02$sedentary, adult02$bmi_t)
xt_2x3(adult06$sedentary, adult06$bmi_t)
xt_2x3(adult11$sedentary, adult11$bmi_t)
xt_2x3(adult12$sedentary, adult12$bmi_t)
xt_2x3(adult13$sedentary, adult13$bmi_t)
xt_2x3(adult14$sedentary, adult14$bmi_t)

rbind(chix_table(adult02$sedentary, adult02$bmi_t),
      chix_table(adult06$sedentary, adult06$bmi_t),
      chix_table(adult11$sedentary, adult11$bmi_t),
      chix_table(adult12$sedentary, adult12$bmi_t),
      chix_table(adult13$sedentary, adult13$bmi_t),
      chix_table(adult14$sedentary, adult14$bmi_t))

grid.arrange(stacked_plot(adult02, "sedentary", "bmi_t", "Sedentary Lifestyle", "BMI Tertile", "2002
    /03"),
            stacked_plot(adult06, "sedentary", "bmi_t", "Sedentary Lifestyle", "BMI Tertile", "2006/
                07"),
            stacked_plot(adult11, "sedentary", "bmi_t", "Sedentary Lifestyle", "BMI Tertile", "2011/
                12"),
            stacked_plot(adult12, "sedentary", "bmi_t", "Sedentary Lifestyle", "BMI Tertile", "2012/
                13"),
            stacked_plot(adult13, "sedentary", "bmi_t", "Sedentary Lifestyle", "BMI Tertile", "2013/
                14"),
            stacked_plot(adult14, "sedentary", "bmi_t", "Sedentary Lifestyle", "BMI Tertile", "2014/
                15"),
            ncol= 3, nrow=2, padding=unit(3,"line"))


## CHILD DATA
## Subset the child data on each year
child06 = subset(child, year == "2006/07")
child11 = subset(child, year == "2011/12")
child12 = subset(child, year == "2012/13")
child13 = subset(child, year == "2013/14")
child14 = subset(child, year == "2014/15")
```

```
# AGE
grid.arrange(box_plot_median(child06, "bmic", "age", "Age", "BMI Category", "2006/07", max= 20, bins
    =5, txt.size = 4),
             box_plot_median(child11, "bmic", "age", "Age", "BMI Category", "2011/12", max= 20, bins
                 =5, txt.size = 4),
             box_plot_median(child12, "bmic", "age", "Age", "BMI Category", "2012/13", max= 20, bins
                 =5, txt.size = 4),
             box_plot_median(child13, "bmic", "age", "Age", "BMI Category", "2013/14", max= 20, bins
                 =5, txt.size = 4),
             box_plot_median(child14, "bmic", "age", "Age", "BMI Category", "2014/15", max= 20, bins
                 =5, txt.size = 4),
             ncol=3, nrow=2, padding = unit(3, "line"))


rbind(anova_table(child06$age, child06$bmic),
anova_table(child11$age, child11$bmic),
anova_table(child12$age, child12$bmic),
anova_table(child13$age, child13$bmic),
anova_table(child14$age, child14$bmic))



# SEX
xt_2x3(child06$gender, child06$bmic)
xt_2x3(child11$gender, child11$bmic)
xt_2x3(child12$gender, child12$bmic)
xt_2x3(child13$gender, child13$bmic)
xt_2x3(child14$gender, child14$bmic)

rbind(chix_table(child06$gender, child06$bmic),
      chix_table(child11$gender, child11$bmic),
      chix_table(child12$gender, child12$bmic),
      chix_table(child13$gender, child13$bmic),
      chix_table(child14$gender, child14$bmic))

grid.arrange(stacked_plot(child06, "gender", "bmic", "Sex Group", "BMI Category", "2006/07", "2--<18
     yo"),
             stacked_plot(child11, "gender", "bmic", "Sex Group", "BMI Category", "2011/12", "2--<18
```

```
yo"),
            stacked_plot(child12, "gender", "bmic", "Sex Group", "BMI Category", "2012/13", "2--<18 yo"),
            stacked_plot(child13, "gender", "bmic", "Sex Group", "BMI Category", "2013/14", "2--<18 yo"),
            stacked_plot(child14, "gender", "bmic", "Sex Group", "BMI Category", "2014/15", "2--<18 yo"),
            ncol=3, nrow=2, padding= unit(3, "line"))


# ETHNICITY
table(child06$eth_count, child06$bmic)
table(child11$eth_count, child11$bmic)
table(child12$eth_count, child12$bmic)
table(child13$eth_count, child13$bmic)
table(child14$eth_count, child14$bmic)

rbind(chix_table(child06$eth_count, child06$bmic),
      chix_table(child11$eth_count, child11$bmic),
      chix_table(child12$eth_count, child12$bmic),
      chix_table(child13$eth_count, child13$bmic),
      chix_table(child14$eth_count, child14$bmic))

grid.arrange(stacked_plot(child06, "eth_count", "bmic", "Ethnic Groups", "BMI Category", "2006/07", "2--<18 yo"),
            stacked_plot(child11, "eth_count", "bmic", "Ethnic Groups", "BMI Category", "2011/12", "2--<18 yo"),
            stacked_plot(child12, "eth_count", "bmic", "Ethnic Groups", "BMI Category", "2012/13", "2--<18 yo"),
            stacked_plot(child13, "eth_count", "bmic", "Ethnic Groups", "BMI Category", "2013/14", "2--<18 yo"),
            stacked_plot(child14, "eth_count", "bmic", "Ethnic Groups", "BMI Category", "2014/15", "2--<18 yo"),
            ncol=3, nrow=2, padding= unit(3, "line"))


# DEPRIVATION
xt_5x3(child06$dep, child06$bmic)
xt_5x3(child11$dep, child11$bmic)
xt_5x3(child12$dep, child12$bmic)
```

```
xt_5x3(child13$dep, child13$bmic)
xt_5x3(child14$dep, child14$bmic)

rbind(chix_table(child06$dep, child06$bmic),
      chix_table(child11$dep, child11$bmic),
      chix_table(child12$dep, child12$bmic),
      chix_table(child13$dep, child13$bmic),
      chix_table(child14$dep, child14$bmic))

grid.arrange(stacked_plot(child06, "dep", "bmic", "Deprivation Quintiles", "BMI Category", "2006/07"
    , "2--<18 yo"),
             stacked_plot(child11, "dep", "bmic", "Deprivation Quintiles", "BMI Category", "2011/12"
                 , "2--<18 yo"),
             stacked_plot(child12, "dep", "bmic", "Deprivation Quintiles", "BMI Category", "2012/13"
                 , "2--<18 yo"),
             stacked_plot(child13, "dep", "bmic", "Deprivation Quintiles", "BMI Category", "2013/14"
                 , "2--<18 yo"),
             stacked_plot(child14, "dep", "bmic", "Deprivation Quintiles", "BMI Category", "2014/15"
                 , "2--<18 yo"),
             ncol=3, nrow=2, padding= unit(3, "line"))


# URBAN/RURAL AREA
xt_2x3(child06$UA, child06$bmic)
xt_2x3(child14$UA, child14$bmic)

rbind(chix_table(child06$UA, child06$bmic),
      chix_table(child14$UA, child14$bmic))

grid.arrange(stacked_plot(child06, "UA", "bmic", "Urban/Rural Area", "BMI Category", "2006/07", "
    2--<18 yo"),
             stacked_plot(child14, "UA", "bmic", "Urban/Rural Area", "BMI Category", "2014/15", "
                 2--<18 yo"),
             ncol=3, nrow=2, padding= unit(3, "line"))


# HOUSEHOLD INCOME
xt_4x3(child06$hhinc, child06$bmic)
xt_4x3(child11$hhinc, child11$bmic)
xt_4x3(child12$hhinc, child12$bmic)
```

```r
xt_4x3(child13$hhinc, child13$bmic)
xt_4x3(child14$hhinc, child14$bmic)

rbind(chix_table(child06$hhinc, child06$bmic),
      chix_table(child11$hhinc, child11$bmic),
      chix_table(child12$hhinc, child12$bmic),
      chix_table(child13$hhinc, child13$bmic),
      chix_table(child14$hhinc, child14$bmic))

grid.arrange(stacked_plot(child06, "hhinc", "bmic", "Household Income", "BMI Category", "2006/07", "
    2−−<18 yo"),
             stacked_plot(child11, "hhinc", "bmic", "Household Income", "BMI Category", "2011/12", "
                2−−<18 yo"),
             stacked_plot(child12, "hhinc", "bmic", "Household Income", "BMI Category", "2012/13", "
                2−−<18 yo"),
             stacked_plot(child13, "hhinc", "bmic", "Household Income", "BMI Category", "2013/14", "
                2−−<18 yo"),
             stacked_plot(child14, "hhinc", "bmic", "Household Income", "BMI Category", "2014/15", "
                2−−<18 yo"),
             ncol=3, nrow=2, padding= unit(3, "line"))



# EDUCATION
xt_3x3(child06$edu, child06$bmic)
xt_3x3(child11$edu, child11$bmic)
xt_3x3(child12$edu, child12$bmic)
xt_3x3(child13$edu, child13$bmic)
xt_3x3(child14$edu, child14$bmic)

rbind(chix_table(child06$edu, child06$bmic),
      chix_table(child11$edu, child11$bmic),
      chix_table(child12$edu, child12$bmic),
      chix_table(child13$edu, child13$bmic),
      chix_table(child14$edu, child14$bmic))

grid.arrange(stacked_plot(child06, "edu", "bmic", "Educational Qualification", "BMI Category", "2006
    /07", "2−−<18 yo"),
             stacked_plot(child11, "edu", "bmic", "Educational Qualification", "BMI Category", "2011
                /12", "2−−<18 yo"),
```

```
             stacked_plot(child12, "edu", "bmic", "Educational Qualification", "BMI Category", "2012/13", "2--<18 yo"),
             stacked_plot(child13, "edu", "bmic", "Educational Qualification", "BMI Category", "2013/14", "2--<18 yo"),
             stacked_plot(child14, "edu", "bmic", "Educational Qualification", "BMI Category", "2014/15", "2--<18 yo"),
             ncol=3, nrow=2, padding= unit(3, "line"))



# FRUIT GUIDELINE
xt_2x3(child06$fruit, child06$bmic)
xt_2x3(child11$fruit, child11$bmic)
xt_2x3(child12$fruit, child12$bmic)
xt_2x3(child13$fruit, child13$bmic)
xt_2x3(child14$fruit, child14$bmic)

rbind(chix_table(child06$fruit, child06$bmic),
      chix_table(child11$fruit, child11$bmic),
      chix_table(child12$fruit, child12$bmic),
      chix_table(child13$fruit, child13$bmic),
      chix_table(child14$fruit, child14$bmic))

grid.arrange(stacked_plot(child06, "fruit", "bmic", "Fruit Guideline", "BMI Category", "2006/07", "2--<18 yo"),
             stacked_plot(child11, "fruit", "bmic", "Fruit Guideline", "BMI Category", "2011/12", "2--<18 yo"),
             stacked_plot(child12, "fruit", "bmic", "Fruit Guideline", "BMI Category", "2012/13", "2--<18 yo"),
             stacked_plot(child13, "fruit", "bmic", "Fruit Guideline", "BMI Category", "2013/14", "2--<18 yo"),
             stacked_plot(child14, "fruit", "bmic", "Fruit Guideline", "BMI Category", "2014/15", "2--<18 yo"),
             ncol=3, nrow=2, padding= unit(3, "line"))



# VEGETABLE GUIDELINE
xt_2x3(child06$veges, child06$bmic)
xt_2x3(child11$veges, child11$bmic)
```

```
xt_2x3(child12$veges, child12$bmic)
xt_2x3(child13$veges, child13$bmic)
xt_2x3(child14$veges, child14$bmic)

rbind(chix_table(child06$veges, child06$bmic),
      chix_table(child11$veges, child11$bmic),
      chix_table(child12$veges, child12$bmic),
      chix_table(child13$veges, child13$bmic),
      chix_table(child14$veges, child14$bmic))

grid.arrange(stacked_plot(child06, "veges", "bmic", "Vegetable Guideline", "BMI Category", "2006/07"
    , "2--<18 yo"),
             stacked_plot(child11, "veges", "bmic", "Vegetable Guideline", "BMI Category", "2011/12"
                 , "2--<18 yo"),
             stacked_plot(child12, "veges", "bmic", "Vegetable Guideline", "BMI Category", "2012/13"
                 , "2--<18 yo"),
             stacked_plot(child13, "veges", "bmic", "Vegetable Guideline", "BMI Category", "2013/14"
                 , "2--<18 yo"),
             stacked_plot(child14, "veges", "bmic", "Vegetable Guideline", "BMI Category", "2014/15"
                 , "2--<18 yo"),
             ncol=3, nrow=2, padding= unit(3, "line"))



# SOFT DRINK CONSUMPTION
xt_4x3(child06$softd, child06$bmic)
xt_4x3(child11$softd, child11$bmic)
xt_4x3(child12$softd, child12$bmic)
xt_4x3(child13$softd, child13$bmic)
xt_4x3(child14$softd, child14$bmic)

rbind(chix_table(child06$softd, child06$bmic),
      chix_table(child11$softd, child11$bmic),
      chix_table(child12$softd, child12$bmic),
      chix_table(child13$softd, child13$bmic),
      chix_table(child14$softd, child14$bmic))

grid.arrange(stacked_plot(child06, "softd", "bmic", "Soft Drink Consumption", "BMI Category", "2006/
    07", "2--<18 yo"),
             stacked_plot(child11, "softd", "bmic", "Soft Drink Consumption", "BMI Category", "2011/
```

```
        12", "2−−<18 yo"),
        stacked_plot(child12, "softd", "bmic", "Soft Drink Consumption", "BMI Category", "2012/
            13", "2−−<18 yo"),
        stacked_plot(child13, "softd", "bmic", "Soft Drink Consumption", "BMI Category", "2013/
            14", "2−−<18 yo"),
        stacked_plot(child14, "softd", "bmic", "Soft Drink Consumption", "BMI Category", "2014/
            15", "2−−<18 yo"),
        ncol=3, nrow=2, padding= unit(3, "line"))


# FAST FOOD CONSUMPTION
xt_4x3(child06$fastf, child06$bmic)
xt_4x3(child11$fastf, child11$bmic)
xt_4x3(child12$fastf, child12$bmic)
xt_4x3(child13$fastf, child13$bmic)
xt_4x3(child14$fastf, child14$bmic)

rbind(chix_table(child06$fastf, child06$bmic),
      chix_table(child11$fastf, child11$bmic),
      chix_table(child12$fastf, child12$bmic),
      chix_table(child13$fastf, child13$bmic),
      chix_table(child14$fastf, child14$bmic))

grid.arrange(stacked_plot(child06, "fastf", "bmic", "Fast Food Consumption", "BMI Category", "2006/
    07", "2−−<18 yo"),
        stacked_plot(child11, "fastf", "bmic", "Fast Food Consumption", "BMI Category", "2011/
            12", "2−−<18 yo"),
        stacked_plot(child12, "fastf", "bmic", "Fast Food Consumption", "BMI Category", "2012/
            13", "2−−<18 yo"),
        stacked_plot(child13, "fastf", "bmic", "Fast Food Consumption", "BMI Category", "2013/
            14", "2−−<18 yo"),
        stacked_plot(child14, "fastf", "bmic", "Fast Food Consumption", "BMI Category", "2014/
            15", "2−−<18 yo"),
        ncol=3, nrow=2, padding= unit(3, "line"))
```

Listing A.4: Bivariate Analyses - Deprivation Quintiles

```
#### BIVARIATE ANALYSES
## Examining confounding factors on the deprivation quintile variable


## Adult Data
# Age
rbind(anova_table(adult02$age, adult02$dep),
      anova_table(adult06$age, adult06$dep),
      anova_table(adult11$age, adult11$dep),
      anova_table(adult12$age, adult12$dep),
      anova_table(adult13$age, adult13$dep),
      anova_table(adult14$age, adult14$dep))


# Sex
rbind(chix_table(adult02$gender, adult02$dep),
      chix_table(adult06$gender, adult06$dep),
      chix_table(adult11$gender, adult11$dep),
      chix_table(adult12$gender, adult12$dep),
      chix_table(adult13$gender, adult13$dep),
      chix_table(adult14$gender, adult14$dep))


# Urban/Rural area
rbind(chix_table(adult02$UA, adult02$dep),
      chix_table(adult06$UA, adult06$dep),
      chix_table(adult14$UA, adult14$dep))


# Household Income
rbind(chix_table(adult02$hhinc, adult02$dep),
      chix_table(adult06$hhinc, adult06$dep),
      chix_table(adult11$hhinc, adult11$dep),
      chix_table(adult12$hhinc, adult12$dep),
      chix_table(adult13$hhinc, adult13$dep),
      chix_table(adult14$hhinc, adult14$dep))
```

```r
# Education
rbind(chix_table(adult02$edu, adult02$dep),
      chix_table(adult06$edu, adult06$dep),
      chix_table(adult11$edu, adult11$dep),
      chix_table(adult12$edu, adult12$dep),
      chix_table(adult13$edu, adult13$dep),
      chix_table(adult14$edu, adult14$dep))


# Fruit Guideline
rbind(chix_table(adult02$fruit, adult02$dep),
      chix_table(adult06$fruit, adult06$dep),
      chix_table(adult11$fruit, adult11$dep),
      chix_table(adult12$fruit, adult12$dep),
      chix_table(adult13$fruit, adult13$dep),
      chix_table(adult14$fruit, adult14$dep))


# Vegetable GUideline
rbind(chix_table(adult02$veges, adult02$dep),
      chix_table(adult06$veges, adult06$dep),
      chix_table(adult11$veges, adult11$dep),
      chix_table(adult12$veges, adult12$dep),
      chix_table(adult13$veges, adult13$dep),
      chix_table(adult14$veges, adult14$dep))


# Migration Status
rbind(chix_table(adult02$native, adult02$dep),
      chix_table(adult06$native, adult06$dep),
      chix_table(adult11$native, adult11$dep),
      chix_table(adult12$native, adult12$dep),
      chix_table(adult13$native, adult13$dep),
      chix_table(adult14$native, adult14$dep))


# Difficulty Climbing Stairs
rbind(chix_table(adult02$stair, adult02$dep),
      chix_table(adult06$stair, adult06$dep),
      chix_table(adult11$stair, adult11$dep),
```

```
      chix_table(adult12$stair, adult12$dep),
      chix_table(adult13$stair, adult13$dep),
      chix_table(adult14$stair, adult14$dep))


# Smoking Status
rbind(chix_table(adult02$smoke, adult02$dep),
      chix_table(adult06$smoke, adult06$dep),
      chix_table(adult11$smoke, adult11$dep),
      chix_table(adult12$smoke, adult12$dep),
      chix_table(adult13$smoke, adult13$dep),
      chix_table(adult14$smoke, adult14$dep))


# Alcohol Problem
rbind(chix_table(adult02$haz_drinker_all, adult02$dep),
      chix_table(adult06$haz_drinker_all, adult06$dep),
      chix_table(adult11$haz_drinker_all, adult11$dep),
      chix_table(adult12$haz_drinker_all, adult12$dep),
      chix_table(adult13$haz_drinker_all, adult13$dep),
      chix_table(adult14$haz_drinker_all, adult14$dep))


# Physical Acitivity
rbind(chix_table(adult02$active, adult02$dep),
      chix_table(adult06$active, adult06$dep),
      chix_table(adult11$active, adult11$dep),
      chix_table(adult12$active, adult12$dep),
      chix_table(adult13$active, adult13$dep),
      chix_table(adult14$active, adult14$dep))


# Sedentary Lifestyle
rbind(chix_table(adult02$sedentary, adult02$dep),
      chix_table(adult06$sedentary, adult06$dep),
      chix_table(adult11$sedentary, adult11$dep),
      chix_table(adult12$sedentary, adult12$dep),
      chix_table(adult13$sedentary, adult13$dep),
      chix_table(adult14$sedentary, adult14$dep))

```

```r
## Child Data
# Age
rbind(anova_table(child06$age, child06$dep),
      anova_table(child11$age, child11$dep),
      anova_table(child12$age, child12$dep),
      anova_table(child13$age, child13$dep),
      anova_table(child14$age, child14$dep))


# Sex
rbind(chix_table(child06$gender, child06$dep),
      chix_table(child11$gender, child11$dep),
      chix_table(child12$gender, child12$dep),
      chix_table(child13$gender, child13$dep),
      chix_table(child14$gender, child14$dep))


# Urban/Rural area
rbind(chix_table(child06$UA, child06$dep),
      chix_table(child14$UA, child14$dep))


# Household Income
rbind(chix_table(child06$hhinc, child06$dep),
      chix_table(child11$hhinc, child11$dep),
      chix_table(child12$hhinc, child12$dep),
      chix_table(child13$hhinc, child13$dep),
      chix_table(child14$hhinc, child14$dep))


# Education
rbind(chix_table(child06$edu, child06$dep),
      chix_table(child11$edu, child11$dep),
      chix_table(child12$edu, child12$dep),
      chix_table(child13$edu, child13$dep),
      chix_table(child14$edu, child14$dep))
```

```r
# softd Guideline
rbind(chix_table(child06$fruit , child06$dep),
      chix_table(child11$fruit , child11$dep),
      chix_table(child12$fruit , child12$dep),
      chix_table(child13$fruit , child13$dep),
      chix_table(child14$fruit , child14$dep))


# Vegetable GUideline
rbind(chix_table(child06$veges, child06$dep),
      chix_table(child11$veges, child11$dep),
      chix_table(child12$veges, child12$dep),
      chix_table(child13$veges, child13$dep),
      chix_table(child14$veges, child14$dep))


# Soft Drink Consumption
rbind(chix_table(child06$softd , child06$dep),
      chix_table(child11$softd , child11$dep),
      chix_table(child12$softd , child12$dep),
      chix_table(child13$softd , child13$dep),
      chix_table(child14$softd , child14$dep))


# Fast Food Consumption
rbind(chix_table(child06$fastf , child06$dep),
      chix_table(child11$fastf , child11$dep),
      chix_table(child12$fastf , child12$dep),
      chix_table(child13$fastf , child13$dep),
      chix_table(child14$fastf , child14$dep))
```

B Bivariate Tables

B.1 Bivariate Tables for The Adult Data

Table B.1: Mean Age by BMI Tertiles

BMI Tertiles	Age (in years) mean (SD)
Bottom Tertile	46.08 (18.92)
Mid Tertile	50.36 (17.83)
Top Tertile	49.29 (16.45)

Table B.2: Sex by BMI Tertiles

BMI Tertiles	Male	Female
	N (row %)	
Bottom Tertile	8570 (0.38)	14007 (0.62)
Mid Tertile	11539 (0.51)	11038 (0.49)
Top Tertile	9237 (0.41)	13340 (0.59)

Table B.3: Ethnicities by BMI Tertiles

BMI Tertiles	European Only	2+ Ethnicities (M)	Asian Only	Maori Only	Pacific Only	Other
	N (row %)					
Bottom Tertile	13645 (0.61)	1619 (0.07)	1997 (0.09)	1783 (0.08)	246 (0.01)	3230 (0.14)
Mid Tertile	13719 (0.61)	1855 (0.08)	1058 (0.05)	2714 (0.12)	715 (0.03)	2452 (0.11)
Top Tertile	10534 (0.47)	2337 (0.10)	373 (0.02)	5167 (0.23)	2320 (0.10)	1804 (0.08)

Table B.4: Deprivation Quintiles by BMI Tertiles

BMI Tertiles	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
	N (row %)				
Bottom Tertile	3779 (0.17)	4058 (0.18)	4688 (0.21)	5043 (0.22)	5006 (0.22)
Mid Tertile	3507 (0.16)	3950 (0.18)	4414 (0.20)	5053 (0.22)	5640 (0.25)
Top Tertile	2233 (0.10)	2693 (0.12)	3781 (0.17)	5181 (0.23)	8678 (0.38)

Table B.5: Urban/Rural Area by BMI Tertiles

BMI Tertiles	Urban	Rural
	N (row %)	
Bottom Tertile	9563 (0.83)	1910 (0.17)
Mid Tertile	9124 (0.81)	2175 (0.19)
Top Tertile	8868 (0.79)	2367 (0.21)

Table B.6: Household Income by BMI Tertiles

BMI Tertiles	≤$15,000	$15,001-$40,000	$40,001–$70,000	>$70,000
	N (row %)			
Bottom Tertile	1452 (0.08)	5280 (0.31)	4325 (0.25)	6106 (0.36)
Mid Tertile	1362 (0.08)	5478 (0.31)	4475 (0.25)	6639 (0.37)
Top Tertile	1716 (0.10)	5942 (0.34)	4273 (0.25)	5442 (0.31)

Table B.7: Educational Qualifications by BMI Tertiles

BMI Tertiles	No Qualification	Secondary	Tertiary
	N (row %)		
Bottom Tertile	3623 (0.16)	6866 (0.31)	11923 (0.53)
Mid Tertile	4576 (0.20)	6487 (0.29)	11363 (0.51)
Top Tertile	5988 (0.27)	7044 (0.32)	9319 (0.42)

Table B.8: Adherence to Fruit Guideline by BMI Tertiles

BMI Tertiles	No	Yes
	N (row %)	
Bottom Tertile	9610 (0.43)	12956 (0.57)
Mid Tertile	9762 (0.43)	12807 (0.57)
Top Tertile	10298 (0.46)	12267 (0.54)

Table B.9: Adherence to Vegetable Guideline by BMI Tertiles

BMI Tertiles	No	Yes
	N (row %)	
Bottom Tertile	8892 (0.39)	13674 (0.61)
Mid Tertile	8609 (0.38)	13956 (0.62)
Top Tertile	9139 (0.41)	13422 (0.59)

Table B.10: Migration Status by BMI Tertiles

BMI Tertiles	Migrant	Native
	N (row %)	
Bottom Tertile	2576 (0.11)	19990 (0.89)
Mid Tertile	1489 (0.07)	21077 (0.93)
Top Tertile	985 (0.04)	21574 (0.96)

Table B.11: Difficulty Climbing Stairs by BMI Tertiles

BMI Tertiles	No Difficulty	A Little Difficult	A Lot Difficult
	N (row %)		
Bottom Tertile	18708 (0.83)	2297 (0.10)	1522 (0.07)
Mid Tertile	17718 (0.79)	2930 (0.13)	1888 (0.08)
Top Tertile	15251 (0.68)	4165 (0.18)	3121 (0.14)

Table B.12: Smoking Status by BMI Tertiles

BMI Tertiles	Non Smoker	Ex Smoker	Current Smoker
	N (row %)		
Bottom Tertile	10296 (0.55)	4201 (0.22)	4260 (0.23)
Mid Tertile	9395 (0.50)	5626 (0.30)	3949 (0.21)
Top Tertile	8483 (0.44)	6127 (0.32)	4475 (0.23)

Table B.13: Drinking Problem by BMI Tertiles

BMI Tertiles	No Drinking Problem	Drinking Problem
	N (row %)	
Bottom Tertile	18879 (0.84)	3538 (0.16)
Mid Tertile	18455 (0.82)	3947 (0.18)
Top Tertile	18541 (0.83)	3842 (0.17)

Table B.14: Physical Activity by BMI Tertiles

BMI Tertiles	Physically Active	Not Active
	N (row %)	
Bottom Tertile	12251 (0.55)	10163 (0.45)
Mid Tertile	12023 (0.54)	10379 (0.46)
Top Tertile	10667 (0.48)	11735 (0.52)

Table B.15: Sedentary Lifestyle by BMI Tertiles

BMI Tertiles	Not Sedentary	Sedentary
	N (row %)	
Bottom Tertile	19378 (0.86)	3043 (0.14)
Mid Tertile	19321 (0.86)	3092 (0.14)
Top Tertile	18297 (0.82)	4114 (0.18)

B.2 Bivariate Tables for The Child Data

Table B.16: Mean Age by BMI Categories

BMI Categories	Age (in years)
	mean (SD)
Average	8.58 (4.40)
Overweight	9.29 (4.30)
Obese	9.56 (4.26)

Table B.17: Sex by BMI Categories

BMI Categories	Male	Female
	N (row %)	
Average	6026 (0.53)	5435 (0.47)
Overweight	2300 (0.50)	2346 (0.50)
Obese	1431 (0.50)	1410 (0.50)

Table B.18: Ethnicities by BMI Categories

BMI Categories	European Only	2+ Ethnicities (M)	Asian Only	Maori Only	Pacific Only	Other
	N (row %)					
Average	5041 (0.44)	2288 (0.20)	616 (0.05)	1481 (0.13)	1568 (0.14)	451 (0.04)
Overweight	1544 (0.33)	1075 (0.23)	186 (0.04)	891 (0.19)	527 (0.11)	422 (0.09)
Obese	612 (0.22)	626 (0.22)	79 (0.03)	705 (0.25)	273 (0.10)	545 (0.19)

Table B.19: Deprivation Quintiles by BMI Categories

BMI Categories	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
	N (row %)				
Average	1864 (0.16)	1975 (0.17)	2123 (0.19)	2506 (0.22)	2993 (0.26)
Overweight	537 (0.12)	619 (0.13)	777 (0.17)	1073 (0.23)	1640 (0.35)
Obese	175 (0.06)	245 (0.09)	366 (0.13)	630 (0.22)	1425 (0.50)

Table B.20: Urban/Rural Area by BMI Categories

BMI Categories	Urban	Rural
	N (row %)	
Average	4242 (0.83)	868 (0.17)
Overweight	1735 (0.84)	329 (0.16)
Obese	1077 (0.85)	184 (0.15)

Table B.21: Household Income by BMI Categories

BMI Categories	≤$15,000	$15,001-$40,000	$40,001–$70,000	>$70,000
	N (row %)			
Average	3977 (0.46)	367 (0.04)	1920 (0.22)	2307 (0.27)
Overweight	1355 (0.40)	186 (0.06)	891 (0.27)	915 (0.27)
Obese	621 (0.33)	131 (0.07)	580 (0.31)	535 (0.29)

Table B.22: Parents' Educational Qualifications by BMI Categories

BMI Categories	No Qualification	Secondary	Tertiary
	N (row %)		
Average	1709 (0.15)	4055 (0.37)	5280 (0.48)
Overweight	891 (0.20)	1782 (0.40)	1817 (0.40)
Obese	709 (0.26)	1205 (0.44)	827 (0.30)

Table B.23: Adherence to Fruit Guideline by BMI Categories

BMI Categories	No	Yes
	N (row %)	
Average	2620 (0.29)	6374 (0.71)
Overweight	1160 (0.31)	2533 (0.69)
Obese	789 (0.34)	1558 (0.66)

Table B.24: Adherence to Vegetable Guideline by BMI Categories

BMI Categories	No	Yes
	N (row %)	
Average	3885 (0.43)	5106 (0.57)
Overweight	1576 (0.43)	2114 (0.57)
Obese	1166 (0.50)	1184 (0.50)

Table B.25: Soft Drink Consumption by BMI Categories

BMI Categories	0/week	1/week	2–3/week	4+/week
	N (row %)			
Average	4005 (0.39)	2793 (0.27)	2196 (0.21)	1275 (0.12)
Overweight	1330 (0.32)	1096 (0.27)	1045 (0.25)	639 (0.16)
Obese	682 (0.28)	641 (0.26)	667 (0.27)	445 (0.18)

Table B.26: Fast Food Consumption by BMI Categories

BMI Categories	0/week	1/week	2–3/week	4+/week
	N (row %)			
Average	3161 (0.31)	5010 (0.49)	1856 (0.18)	257 (0.02)
Overweight	1148 (0.28)	1950 (0.47)	882 (0.21)	142 (0.03)
Obese	595 (0.24)	1103 (0.45)	632 (0.26)	113 (0.05)

B.3 Summary of Bivariate Statistical Tests

Table B.27: Bivariate Analysis of The Covariates and BMI Tertiles (Adult Data)

Covariates	*F*	df (between, within)	p	η^2	Effect Size
Age	355.25	2, 67728	<.001	.010	*

Covariates	χ^2	df	p	φ_c	Effect Size
Sex	875.59	2	<.001	.114	*
Ethnicity	6292.52	10	<.001	.216	**
Deprivation Quintile	2051.51	8	<.001	.123	*
Urban/Rural Area	72.99	2	<.001	.023	
Household Income	190.30	6	<.001	.038	
Education	969.21	4	<.001	.085	*
Fruit Guideline	47.09	2	<.001	.019	
Vegetable Guideline	26.27	2	<.001	.014	
Migration Status	848.29	2	<.001	.079	
Difficulty Climbing Several Flights of Stairs	1589.52	4	<.001	.108	*
Smoking Status	581.53	4	<.001	.066	
Drinking Problem	29.28	2	<.001	.015	
Physical Activity	261.00	2	<.001	.044	
Sedentary Lifestyle	253.08	2	<.001	.043	

F = F-statistic from ANOVA; χ^2 = Chi-squared statistic
η^2 = Eta-squared; φ_c = Cramer's V
* small effect size; ** medium effect size

Table B.28: Bivariate Analysis of The Covariates and BMI Categories (Child Data)

Covariates	*F*	df (between, within)	p	η^2	Effect Size
Age	82.44	2, 18945	<.001	.009	*

Covariates	χ^2	df	p	φ_c	Effect Size
Sex	14.05	2	<.001	.026	
Ethnicity	1390.14	10	<.001	.149	**
Deprivation Quintile	793.49	8	<.001	.112	*
Urban/Rural Area	4.61	2	.099	.011	
Household Income	150.86	6	<.001	.049	
Education	341.58	4	<.001	.074	*
Fruit Guideline	20.23	2	<.001	.022	
Vegetable Guideline	34.99	2	<.001	.029	
Soft Drink Consumption	183.59	6	<.001	.054	
Fast Food Consmption	134.29	6	<.001	.046	

F = F-statistic from ANOVA; χ^2 = Chi-squared statistic
η^2 = Eta-squared; φ_c = Cramer's V
* small effect size; ** medium effect size

C Proportional Odds Assumption

Table C.1: Odds Ratios from Binomial Regressions with Contrasts in The Adult Data

	$\frac{P(Mid\, Tertile\ or\ Top\, Tertile)}{P(Bottom\, Tertile)}$	$\frac{P(Top\, Tertile)}{P(Bottom\, Tertile\ or\ Mid\, Tertile)}$
	OR (95% CI)	
Deprivation Quintile (ref. group: Quintile 1)		
Quintile 2	1.08 (1.03–1.13)	0.94 (0.89–1.00)
Quintile 3	1.09 (1.04–1.15)	1.10 (1.05–1.16)
Quintile 4	1.25 (1.19–1.31)	1.33 (1.26–1.40)
Quintile 5	1.42 (1.35–1.50)	1.49 (1.41–1.58)
Linear Contrast	1.12 (1.07–1.17)	1.28 (1.22–1.33)
Ethnicity (ref. group: European only)		
2+ ethnicities (M)	1.71 (1.62–1.82)	1.67 (1.58–1.75)
Asian only	0.58 (0.54–0.62)	0.41 (0.38–0.45)
Māori only	2.96 (2.79–3.14)	2.84 (2.71–2.99)
Pacific only	9.50 (8.30–10.86)	7.50 (6.95–8.10)
Other	0.89 (0.85–0.93)	0.94 (0.89–0.99)
Age	1.02 (1.02–1.02)	1.003 (1.002–1.004)
Difficulty Climbing Several Flights of Stairs (ref. group: No difficulty)		
A little difficulty climbing stairs	1.61 (1.54–1.69)	2.01 (1.92–2.10)
A lot difficulty climbing stairs	1.56 (1.47–1.66)	2.02 (1.92–2.14)
Sex		
Female*	0.60 (0.58–0.61)	1.04 (1.01–1.08)
Smoking Status (ref. group: Non smoker)		
Ex smoker	1.41 (1.36–1.46)	1.34 (1.30–1.39)
Current smoker	0.92 (0.89–0.96)	0.92 (0.88–0.96)
Household Income (ref. group: $\leq$\$15,000)		
\$15,001–\$40,000	1.18 (1.11–1.25)	1.22 (1.15–1.30)
\$40,001–\$70,000	1.46 (1.37–1.56)	1.37 (1.28–1.46)
>\$70,000	1.69 (1.58–1.80)	1.49 (1.40–1.59)
Educational Qualification (ref. group: No qualification		
Secondary	0.92 (0.88–0.96)	0.87 (0.83–0.91)
Tertiary	0.83 (0.79–0.87)	0.76 (0.73–0.79)
Physical Activity		
Not physically active	1.13 (1.09–1.16)	1.17 (1.14–1.21)
Migration Status		
Native	1.21 (1.14–1.28)	1.20 (1.12–1.29)

* Female had lower odds of being in the top two tertiles compared with being in the bottom tertile, but they had higher odds of being in the top tertile compared with being in the bottom two tertiles

Table C.2: Odds Ratios from Binomial Regressions with Contrasts in The Child Data

	$\frac{P(Overweight\ or\ Obese)}{P(Average)}$	$\frac{P(Obese)}{P(Average\ or\ Overweight)}$
	OR (95%CI)	
Deprivation Quintile		
(ref. group: Quintile 1)		
Quintile 2	1.12 (1.02–1.23)	1.27 (1.09–1.49)
Quintile 3	1.40 (1.28–1.53)	1.70 (1.47–1.96)
Quintile 4	1.49 (1.37–1.63)	1.82 (1.59–2.10)
Quintile 5	1.68 (1.54–1.82)	2.30 (2.01–2.64)
Linear Contrasts	1.15 (1.09–1.21)	1.19 (1.12–1.28)
Ethnicity		
(ref. group: European only)		
2+ ethnicities (M)	1.48 (1.39–1.57)	1.65 (1.51–1.80)
Asian only	1.03 (0.93–1.15)	1.12 (0.94–1.35)
Māori only	1.99 (1.86–2.14)	2.25 (2.05–2.48)
Pacific only	3.48 (3.18–3.81)	4.54 (4.08–5.06)
Other	1.19 (1.11–1.28)	1.33 (1.20–1.48)
Age	1.05 (1.04–1.05)	1.03 (1.02–1.04)
Parents' Qualification		
(ref. group: No qualification)		
Secondary qualification	0.74 (0.69–0.79)	0.74 (0.68–0.80)
Tertiary qualification	0.64 (0.60–0.69)	0.58 (0.53–0.63)
Soft Drink Consumption		
(ref. group: <1/week)		
1/week	1.15 (1.09–1.22)	1.22 (1.12–1.33)
2-3x/week	1.22 (1.15–1.30)	1.34 (1.23–1.46)
4+week	1.18 (1.10–1.26)	1.40 (1.27–1.54)
Sex		
Female	1.12 (1.07–1.17)	1.04 (0.98–1.11)

D Codebook

Table D.1: Codebook for Derived and Renamed Variables in This Thesis

Variable Name	Value Description
gender	Male, Female
bmiscale	Body Mass Index in kg/m2
bmic	Average = a person with an average BMI, Overweight = an overweight person, Obese = an obese person
age	Age in years
euro	European ethnicity; 1 = Yes, 2 = No
maori	Māori ethncity; 1 = Yes, 2 = No
pacific	Pacific ethnicity; 1 = Yes, 2 = No
asian	Indian and/or Chinese ethnicity; 1 = Yes, 2 = No
other	Ethnicity not included elsewhere; 1 = Yes, 2 = No
eth_count	M{a}ori Only = a person who only belong to M{a}ori ethnic group, Pacific Only = a person who only belong to Pacific ethnic group, European Only = a person who only belong to European ethnic group, Asian = a person who belong to Asian and/or Indian ethnic group, 2+ Ethnicities (M) = a person who belong to multiple ethnicities with M{a}ori ancestry, Other = a person who belong to multiple ethnicities or other ethnicities not elsewhere included.
native	Migrant = a person who moved to settle in NZ less than 10-11 years ago, Native = New Zealanders and those who moved to settle in NZ more than 10-11 years ago
secondary	1 = no qualification, 2 = secondary qualification
tertiary	0 = no tertiary qualification, 3 = tertiary qualification
edu	No Qualification = no educational qualification, Secondary = secondary educational qualification (i.e. National Certificate level 1 to 2, NZ University Entrance before 1986 in one or more subhects, NZ Higher School Certificate, University Entrance qualification, NZ A or B Bursary; Scholarship; or National Certificate Level 3, Other NZ secondary school qualification, and Overseas secondary school qualification), Tertiary = tertiary educational qualification (i.e. Bachelor's degree, Bachelors degree with honours, MA, Msc, PhD, Diploma, Diploma - Post Graduate, Trade or technical certificate which took more than 3 months full time study, Professional qualifications, and other tertiary qualifications)
hhinc	Household income per year: <=$15,000; $15,001–$40,001–$70,000; >$70,000
dep	NZ deprivation quintile; 1 = quintile 1, 2 = quintile 2, 3 = quintile 3, 4 = quintile 4, 5 = quintile 5; 2002/03 NZHS used deprivation quintile data from the 2002 Census, 2006/07–2012/13 NZHS used deprivation quintile data from the 2006 Census, 2013/14–2014/15 NZHS used deprivation quintile data from the 2013 Census
UA	Urban = an area with at least a 1000 population, Rural = an area with less than 1000 population
haz_drinker_all	Alcohol Problem = AUDIT score >=8, No Alcohol Problem = AUDIT score <8
minutes	duration of physical activity in the past week in minutes (rigorous physical activity is multiplied by two)
active	Active = a person who spent 30 minutes of brisk walk or moderate physical exercise (15 minutes if it was a vigorous physical activity) in the past week, Not Active = a person who spent less than 30 minutes of brisk walk or moderate physical activity (less than 15 minutes if it was vigorous physical activity) in the past week
sedentary	Sedentary = a person who had not had at least 30 minutes of exercise in the past week, Not Sedentary = a person who had at least 30 minutes of exercise in the past week.
stair	A Lot Difficult = a person who had a lot difficulties climbing several flights of stairs, A Little Difficult = a person who had a little difficulties climbing several flights of stairs, No Difficulty = a person who had no difficulty climbing several flights of stairs
fruit	Yes = a person who met 2+ servings of fruit per day, No = a person who had less than 2 servings of fruit per day
veges	Yes = a person who met 3+ servings of vegetable per day, No = a person who had less than 3 servings of vegetable per day
softd	Soft drink consumption: 0/week, 1/week, 2–3/week or 4+/week
fastf	Fast food consumption: 0/week, 1/week, 2–3/week or 4+/week
cluster	Primary Sampling Unit (PSU) Indicator
strata	Stratum Indicator
finalwgt	Final sample weight